Analysis of Proteomic Alterations during Human Colorectal Tumorigenesis

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Anuli C. Uzozie

aus

Nigeria

Promotionskomitee

Prof. Dr. Josef Jiricny (Vorsitz der Dissertation)

PD. Dr. Giancarlo Marra (Leitung der Dissertation)

Prof. Dr. Paola Picotti

Dr. Paolo Nanni

Zürich, 2014

Table of contents

1 Zusammenfassung

Darmkrebs, insbesondere das kolorektale Karzinom, ist weltweit eine der häufigsten Krebstodesursachen und wird vorwiegend im fortgeschrittenen Alter diagnostiziert. Die nicht erblich bedingte Form von Darmkrebs entsteht meist aus anfangs gutartigen Darmpolypen (Adenomen). Diese entwickeln sich durch molekulare Veränderungen im Darmepithelgewebe, die vorwiegend den Wnt-APC-β-catenin Signalweg betreffen. Adenome gelten als Vorstufe von Krebs und können sich durch zusätzliche genetische und epigenetische Veränderungen zu einem bösartigen Adenokarzinom entwickeln. Eine frühzeitige Entdeckung und Entfernung dieser Krebsvorstufen gilt daher nach wie vor als beste Vorbeugung von Darmkrebs. Screening-Verfahren anhand gastroenterologischer Endoskopie gelten als sehr effizient, sind jedoch invasiv, kosten- und zeitaufwändig (für Personal und Patienten). Nicht zuletzt deswegen stossen endoskopische Verfahren auf wenig Zustimmung seitens der Patienten. Die häufiger verwendeten Tests mittels Stuhlproben sind kostengünstiger und einfacher zu handhaben, haben jedoch im Vergleich zu endoskopischen Untersuchungen eine weit geringere Aussagekraft. Daher ist die Entwicklung einer nichtinvasiven und sensitiven Diagnostikmethode für die Detektion verdächtiger Adenome und früher Adenokarzinome anhand von spezifischen Biomarkern ein grosses Bedürfnis.

Im Zusammenhang mit kolorektalen Adenomen sind zahlreiche genetische und epigenetische Veränderungen dokumentiert worden, jedoch wurden deren funktionelle Auswirkungen auf Proteinebene noch nicht umfassend untersucht. Auf Massenspektrometrie (MS) basierende Studien sind zu einem Eckpfeiler der Krebsproteomforschung geworden und Fortschritte in dieser Technologie haben zu einer deutlichen Verbesserung der Sensitivität und der Reproduzierbarkeit geführt und erlauben eine weitaus umfassendere Proteinanalyse. Bis heute haben sich die meisten Proteomstudien im Zusammenhang mit Darmkrebs auf Veränderungen zwischen Adenokarzinom und Normalgewebe konzentriert; nur wenige Untersuchungen zum Proteinprofil von präinvasiven Adenomen existieren.

Um ein grösseres Verständnis zur Biologie von kolorektalen Adenomen zu erhalten, und um potentielle Marker für die Erkennung früher Darmkrebsstadien zu identifizieren, haben wir umfassende Proteomanalysen von Adenomen mittels quantitativer Shotgun-Massenspektrometrie durchgeführt (Siehe Appendix I). Mittels

iTRAQ 8-plex labelling, OFFGEL Elektrophorese und HPLC- gekoppelter Tandem-Massenspektrometrie haben wir das Proteom von 30 kolorektalen Adenomen und gepaarter Normalschleimhaut untersucht (beide Gewebsproben wurden während einer Koloskopie entnommen). Zusätzlich wurde das Proteom von normalen Epithelzellen (HCEC) und kanzerösen Darmepithelzelllinien (SW480, SW620, CACO2, HT29, CX1) untersucht. Die Resultate aus diesen Analysen zeigen neue Aspekte von klinischer Bedeutung in Bezug auf die untersuchten Krebsvorstufen auf. Insbesondere konnten wir mittels multivariater Analysen zeigen, dass das Proteom der Adenome sich deutlich vom Normalgewebe abgrenzen lässt. Ebenso lassen sich die Krebszelllinien von den HCEC Zellen unterscheiden. Wir haben 212 Proteine identifiziert, die in den Adenomen im Vergleich zum Normalgewebe entweder signifikant hoch- (76) oder herunterreguliert (136) waren. Die meisten identifizierten Änderungen auf Proteinebene entsprachen den Veränderungen auf Transkriptomebene, die wir in unserer Gruppe in bereits in früheren Studien identifizieren konnten. 51 dieser Veränderungen sind sehr wahrscheinlich auf das Epithel der Adenome zurückzuführen, da diese sowohl in den Zelllinien (Krebszellen vs. HCEC) als auch in den Biopsieproben (Adenoma versus Normalgewebe) beobachtet wurden und jeweils ähnliche Veränderungsmuster zeigten.

Auf der Suche nach potentiellen Biomarkern für die Diagnose von Darmkrebs im Frühstadium haben wir die Liste der hochregulierten Proteine in Adenomen mit den verfügbaren Daten der „Human Protein Atlas Database“ verglichen. Diese Datenbank enthält Informationen über Proteinexpression in Krebs- und Normalgewebe. Wir entdeckten, dass bereits Adenome die Proteinveränderungen aufweisen, die in früheren Studien für fortgeschrittene Darmkrebsproben dokumentiert wurden. Ausserdem waren bereits 16 der von uns entdeckten 76 hochregulierten Proteine in dieser Datenbank als Biomarkerkandidaten für Darmkrebs eingetragen.

Eine der herausragendensten Veränderungen auf Ebene der Proteinexpression, die wir in Adenomen gefunden haben, betraf die Hochregulierung der Sorbitol dehydrogenase (SORD), einem Schlüsselenzym im Polyol-Stoffwechselweg. Über die mögliche Rolle von SORD während der Tumorgenese ist in der Literatur nur sehr wenig bekannt. Unsere Validierungsstudien zeigten eine deutlich erhöhte Konzentrationen und Aktivität von SORD in Adenomen und in den Krebszelllinien. Zusätzlich waren Expressionsänderungen anderer Enzyme im gleichen (Aldose Reduktase, AKR1B1) sowie verwandten (Ketohexokinase, KHK) Stoffwechselwegen in Adenomen deutlich dereguliert. Die Nutzung des Polyol-Stoffwechselweges durch Adenomazellen könnte möglicherweise einen Selektionsvorteil gegenüber normalen

Zellen darstellen. Zusätzlich zur Glykolyse und dem Pentosephosphatweg, die beide in Tumoren erhöht sind, könnte der Polyol-Stoffwechsel einen weiteren Weg zur Glukosenutzung in Tumoren darstellen.

Die meisten Proteomanalysestudien über Darmkrebs führten keine systematische Überprüfung und Validierung der gefundenen potentiellen Biomarker in Hinblick auf eine klinische Anwendung durch. Dies ist hauptsächlich dadurch zu erklären, dass robuste Methoden für den genauen und reproduzierbaren Nachweis von gleichzeitig mehreren Tumormarkern und für eine grosse Anzahl von Patientenproben fehlen. Jedoch können neue Methoden auf dem Gebiet der gezielten Proteomanalyse genutzt werden, um konventionelle, immunologische Methoden zu ergänzen und die Validierung möglicher Biomarker zu beschleunigen.

Im zweiten Teil dieser Studie demonstrieren wir die Anwendung von „selected reaction monitoring“ (SRM), einer Technologie zur gezielten Proteomanalyse, um eine Auswahl an möglichen Darmkrebs-Biomarkern in Proben einer grösseren Patientenkohorte zu detektieren und zu quantifizieren. Mit dem Ansatz, diese Untersuchung sowohl in Adenomen wie auch in Karzinomen durchzuführen, demonstrierten wir eine robuste Methode um Veränderungen auf Proteinebene aufzuspüren, die einerseits schon früh in der Darmkrebsentstehung vorhanden sind und andererseits identifizierten wir auch solche, die während der Transformation von Adenomen zu Karzinomen bestehen bleiben.

Proteinbiomarker, die in dieser Arbeit als Kandidaten validiert wurden, entstammen aus unserer vorangegangenen quantitativen (iTRAQ 8plex) „Shotgun-Discovery“-Studie über humane Darmkrebsvorstufen. Die Probenvorbereitung und Analyseprotokolle wurden optimiert, um die Reproduzierbarkeit und Analysepräzision der SRM Messungen sicherzustellen. Verschiedene Softwarepakete für die SRM Assay Entwicklung (Skyline), die automatische Peakgruppenidentifizierung (mProphet) und für Proteinstatistik (MSstats) wurden in einen Arbeitsschritt zusammengefasst. Die SRM Experimente wurden in mehreren Schritten optimiert: Zunächst wurden SRM Koordinaten anhand der Fragmentspektren von Schwerisotop-markierten Referenzpeptiden ermittelt und anschließend durch Messung dieser Referenzpeptide und endogener proteotypischer Peptide in programmierten SRM Sequenzen verfeinert.

Zusätzlich zeigten wir, dass unsere SRM Untersuchungen die reproduzierbare Detektion und die Quantifizierung ausgewählter Proteinmarker in menschlichen Darmproben ermöglichen. Diese Untersuchungen wurden an 72 Gewebeproben (19 Adenomen und 17 Adenokarzinomen mit jeweils benachbartem Normalgewebe) mit 25

ausgewählten Markern durchgeführt. Wir haben eine signifikante Hochregulierung von zehn Proteinen in Darmkrebsvorstufen sowie in Darmkrebs bestätigt (SORD, SPB5, ANXA3, REG4, S10AB, NUCL, NGAL, LDHA, G6PD, AN32A). Gemäss Literatur ist bisher keines dieser Proteine in klinischen Biomarkerstudien validiert worden, obwohl einige der Proteine bereits mit Entstehung von Darmkrebs in Verbindung gebracht worden waren. Alle von uns verifizierten Tumormarker sind potentielle diagnostische Marker für Darmkrebs. Der Workflow mittels SRM bereitet ausserdem eine wertvolle Grundlage für weiterführende Studien, um Darmkrebsbiomarker in anderen Gewebeproben zu verifizieren. Im Moment wenden wir die optimierten SRM Untersuchungen an, um eine Auswahl von Biomarker in den Blutproben einer umfangreichen Kohorte von Darmkrebspatienten zu validieren.

2 Summary

Colorectal cancer is common in elderly individuals and is a major cause of cancer-related deaths worldwide. In the non-hereditary forms, tumorigenesis proceeds along the adenoma-adenocarcinoma pathway. As a result of molecular alterations involving (for the most part) the Wnt-APC-β-catenin signalling pathway, adenomatous tumors arise from the epithelium of the large bowel. These lesions are non-cancerous, but the accumulation over time of additional genetic and epigenetic alterations can transform them into malignant adenocarcinomas. Prompt identification and resection of precancerous lesions is therefore the most effective way to reduce the incidence of colorectal cancer.

Endoscopic screening for these lesions is undeniably effective, but it is also invasive, expensive, and time consuming (for patients and operators). Consequently, a low patient compliance with colorectal cancer screening regulations has been reported worldwide. Faecal-based tests are cheaper and easier to use, but they have lower diagnostic accuracy than colonoscopy-based techniques. There is consequently a vital need for non-invasive, diagnostic tests based on biomarkers with high sensitivity and specificity for high-risk adenomas and early-stage adenocarcinomas.

A multitude of genomic and epigenomic changes have been documented in colorectal adenomas, but their impact on the protein effectors of biological function has not been comprehensively explored. Mass spectrometry (MS) has become the cornerstone of most proteomic studies on cancer development, and ongoing technical advances have markedly improved the sensitivity, reproducibility, and proteome coverage of MS-based studies. Thus far, however, proteomic analyses of human colorectal tissue samples have focused mainly on alterations found in adenocarcinomas, as compared with the normal mucosa. Much less is known about the proteome profile of preinvasive colorectal adenomas.

In an attempt to fill this knowledge gap and to identify promising candidate biomarkers of early-stage colorectal tumorigenesis, we performed a comprehensive proteomic analysis of colorectal adenomas based on quantitative shotgun mass spectrometry (*see Appendix I*). Using iTRAQ 8-plex labelling, OFFGEL electrophoresis, and HPLC-coupled tandem mass spectrometry, we investigated protein expression in 30 colorectal adenomas and paired samples of normal mucosa (prospectively collected

during colonoscopy) and in normal (HCEC) and cancerous (SW480, SW620, CACO2, HT29, CX1) colon epithelial cell lines.

The results revealed several novel aspects of these precancerous lesions with important clinical implications. First of all, in multivariate analysis, the proteomes of adenomas were clearly distinguished from that of normal mucosa, and equally clear distinction emerged between the proteomes of cancer cell lines and HCEC cells. We identified 212 proteins with significantly dysregulated expression in adenomas (upregulated in 76 cases, downregulated in 136) relative to normal mucosal expression. Most of the adenoma-related changes were similar to those previously identified by our group at the transcriptome level. Fifty-one of the changes appear to be specific to the epithelial (rather than stromal) component of adenomas since the proteins in question were all expressed in the colorectal epithelial cell lines, and their altered expression levels in colorectal cancer cells (vs. HCEC cells) were directionally similar to those observed in the adenomas (vs. normal mucosa samples).

Second, to identify potential biomarkers for early-stage colorectal tumorigenesis, we compared the list of proteins exhibiting adenoma-related upregulation with data on protein expression in cancer and normal tissues available in the Human Protein Atlas Database. We discovered that adenomas already exhibit protein expression changes that have been documented in advanced colorectal cancers. Moreover, 16 of the 76 proteins that were upregulated in our adenomas have already been designated as candidate biomarkers for colorectal cancer in this database.

One of the most striking protein expression changes we observed in adenomas involved sorbitol dehydrogenase (SORD), a key enzyme in the polyol pathway whose possible roles in tumorigenesis are largely unexplored. Our validation studies revealed that SORD concentrations and activity are dramatically increased in both colorectal adenomas and cancer cell lines. Other enzymes in the same (aldose reductase, AKR1B1) and related (ketohexokinase, KHK) pathways were also obviously dysregulated in adenomas. Exploitation of the polyol pathway could provide adenomatous cells with a selective advantage over normal cells. Furthermore, the polyol pathway might be another means of tumor-related glucose consumption in addition to the well-known glycolytic and pentose phosphate pathways.

A major shortcoming of most untargeted proteomic studies on cancer is the lack of systematic verification and validation of the identified markers —a fundamental step on the road to their clinical application. Unfortunately, robust methods are lacking for accurate, reproducible verification of multiple tumor markers in a large group of patient

tissue / body fluid samples. However, targeted proteomics methods are now emerging that can complement conventional immunoassay-based techniques and accelerate this verification process.

One example is the method known as selected reaction monitoring (SRM), which we used in the second phase of our study to detect and quantify the abundance of putative colorectal cancer biomarkers in a large cohort of patient tissue samples. Our approach to perform this investigation in both adenomas and adenocarcinomas provided a consistent means to detect protein changes that occur early in colorectal tumorigenesis and those that are likely to persist during the transformation of adenomas to cancerous lesions.

Candidate protein markers verified in this work were previously reported in our quantitative (iTRAQ 8plex) shotgun discovery study on human colorectal precancerous tissues. The sample preparation and analytical protocols were optimized to ensure the reproducibility and analytical precision of all SRM measurements. Software tools for SRM assay development (Skyline), automated peak group identification (mProphet), and protein significance analysis (MSstats) were combined in our comprehensive workflow. Using heavy isotope-labelled reference peptides, we confirmed peptide SRM transitions and refined all transition parameters for SRM measurements.

The SRM assays we developed facilitated the reproducible detection and quantification of candidate biomarker proteins in human colorectal tissues. These assays were applied to monitor 25 proteins in 72 tissues samples comprising neoplastic and adjacent normal mucosa lesions (19 adenoma and normal mucosa pairs; 17 adenocarcinoma and normal mucosa pairs). We confirmed the statistically significant upregulation of ten proteins (SORD, SPB5, ANXA3, REG4, S10AB, NUCL, NGAL, LDHA, G6PD, AN32A) in both precancerous and cancerous colorectal neoplasms (relative to their expression in normal mucosa samples). A review of the literature found that a number of these proteins have been associated/implicated in colorectal tumorigenesis, but none appear to have been validated in clinical biomarker studies. The tumor markers we verified are all potential diagnostic markers for colorectal cancer. Our SRM protocol will be a valuable tool for use in future studies aimed at verifying colorectal biomarkers in other human tissues. We are already exploiting the optimized SRM assays to verify a selected set of putative biomarkers in blood samples from a large number of patients with colorectal cancer.

3 Introduction

The human gastrointestinal (GI) or digestive system is responsible for the breakdown and absorption of nutrients from ingested food/fluids. It consists of specialized organs, beginning from the oral cavity in the mouth, through the stomach and intestines and terminating at the anus, where digested food is passed out as faeces.

The different segments of the GI tract differ in morpholgy, muscular vasculature, enzyme composition and therefore function. Major functions of the GI system include ingestion, digestion (mechanical and chemical), propulsion (peristalsis and segmentation), secretion, absorption and waste elimination. In addition to the main GI organs, accessory organs such as the teeth, salivary glands, tongue, liver, pancreas and gall bladder play specialized roles to maintain the gastrointestinal flora and functions. This is very important as a range of gastrointestinal problems and diseases may occur due to malfunctions in the GI system.

3.1 Anatomy and physiology of the gastrointestinal system (1)

The *upper GI tract* includes digestive organs of the upper abdomen and chest. It comprises of the oral cavity, oesophagus, stomach. The *lower GI tract* includes the small intestine, the large intestine and anus.

The segments of the GI tract are shown in Figure 3.1 (below). The *mouth* is the entrance to the GI tract. It is bounded anteriorly by the teeth and lips, and posteriorly by the oropharynx. The interior surface of the mouth contains the tongue and mylohyoid muscle covered with mucosa. Following the ingestion of food/fluids, enzymes contained in saliva trigger chemical digestion, where as mechanical digestion occurs with the aid of the teeth via mastication.

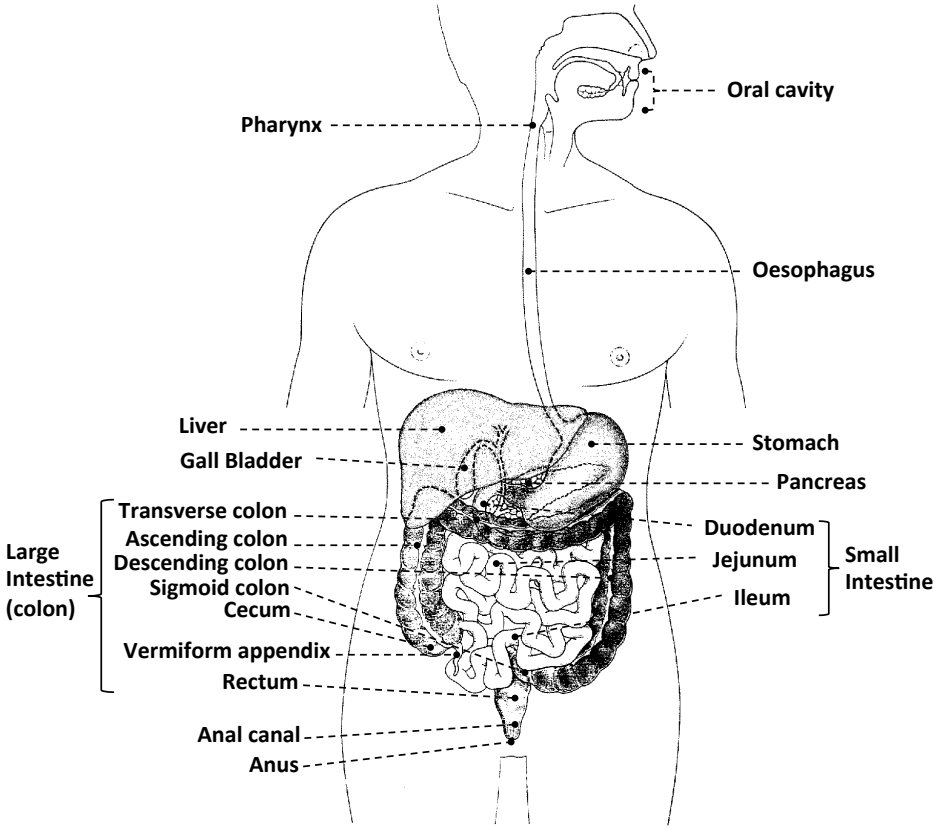


Figure 3.1 The Human Gastrointestinal System

The location of main and accessory organs of the GI tract, with specialized roles for food digestion and nutrient absorption are depicted. Adapted from (2).

Through the action of swallowing, masticated food proceeds through the *pharynx* to the *oesophagus* also know as the gullet. This muscular tube is lined with mucus membranes and facilitates the passage of ingested food from the mouth to the stomach by the process of peristalsis. A sphincter at the end of the oesophagus controls the delivery of food into the stomach and prevents food in the stomach from returning into the oesophagus.

The J-shaped bag at the terminal end of the oesophagus is the *stomach*. It stretches and expands with the entry of food contents from the oesophagus. Ingested food is stored for a short period in the stomach. Churning and mixing motions aid the mechanical breakdown of food. Gastric glands located in the stomach secrete digestive juice and enzymes to enable chemical digestion. Also, stomach acids maintain an acidic environment that is unfavourable to bugs and germs.

The *small intestine* extends from the pylorus of the stomach to the cecum. It is a coiled, approximately 20 feet, thin-walled tube, which is compressed into many folds and occupies a large proportion of the abdominal cavity. The small intestine is

composed of the duodenum, jejunum and ileum. The *duodenum* originates from the pyloric sphincter and forms the first part of the small intestine. Digestive enzymes secreted by the pancreas are delivered into the duodenum as pancreatic juice. A network of thin ducts carry bile produced in the liver and stored in the gall bladder to the duodenum. Bile is a greenish fluid, which aids chemical digestion by breaking down fats into small droplets. These digestive secretions are mixed with partly digested food ejected from the stomach. Chemical digestion and nutrient absorption occurs primarily in the middle segment of the small intestine, also known as the *jejunum*. Nutrients from digested material are absorbed through the wall of the small intestine into the blood stream and lymph. The *ileum* is the last and longest segment of the small intestine. It terminates at the ileocecal valve, a sphincter that monitors the entry of digested materials into the large intestine. Cells in the mucosa of the small intestine have specific functions including mucous secretion, absorption, immune defence, and secretion of digestive enzymes. The hepatic portal system channels blood containing digestive products from the small intestine to the liver.

The *large intestine* has a length of about 5 feet and a diameter of 2.5 inches. It consists of the cecum, colon, rectum and anal canal. Digestive material (chyme) from the ileum is delivered to the *cecum*, a largely peritoneal pouch. It is separated from the ileum by the ileocecal valve and from the colon by the cecocolic junction. Hence, the cecum is often considered as the beginning of the large intestine with direct connection to the first part of the colon. The mammalian *colon* consists of four sections. The *ascending colon* is the first section of the large intestine. It travels up the right side of the abdomen from the cecum. After extraction of water and important nutrients, the digested material is moved upwards to the *transverse colon* by peristalsis. This section of the large intestine runs across the abdomen and is the longest and most flexible part of the colon. Waste material reaching the downward travelling *descending colon*, located at the left abdomen is stored as faeces. The short curving *sigmoid colon* is between the descending colon and the rectum. The contraction of its muscular walls helps deliver faeces to the rectum. The proximal colon comprises the cecum, ascending colon and transverse colon. And the lower or distal colon is made up of the descending colon and sigmoid colon. The human *rectum* is about 4.7 inches long. It begins at the end of the sigmoid colon and faeces released from the sigmoid colon is briefly stored in the rectum. When this occurs, stretch receptors of the nervous system in the rectal walls are stimulated, creating the urge to defaecate. The rectum is followed by the *anal canal*, which terminates at the *anus*. Faeces passes through the

anal canal and is excreted from the anus. Although a fairly low amount of nutrient is absorbed from the digested material that reaches the large intestine, the remaining water and electrolytes are absorbed. A host of beneficial microbial species thrive in the large intestine.

3.2 Histology of the gastrointestinal tract (3-7)

There is a level of consistency in the architectural walls of the digestive tract. As illustrated in Figure 3.2 (below), the walls of the GI tract are made up of four major layers: the serosa, the muscularis, the submucosa and the mucosa.

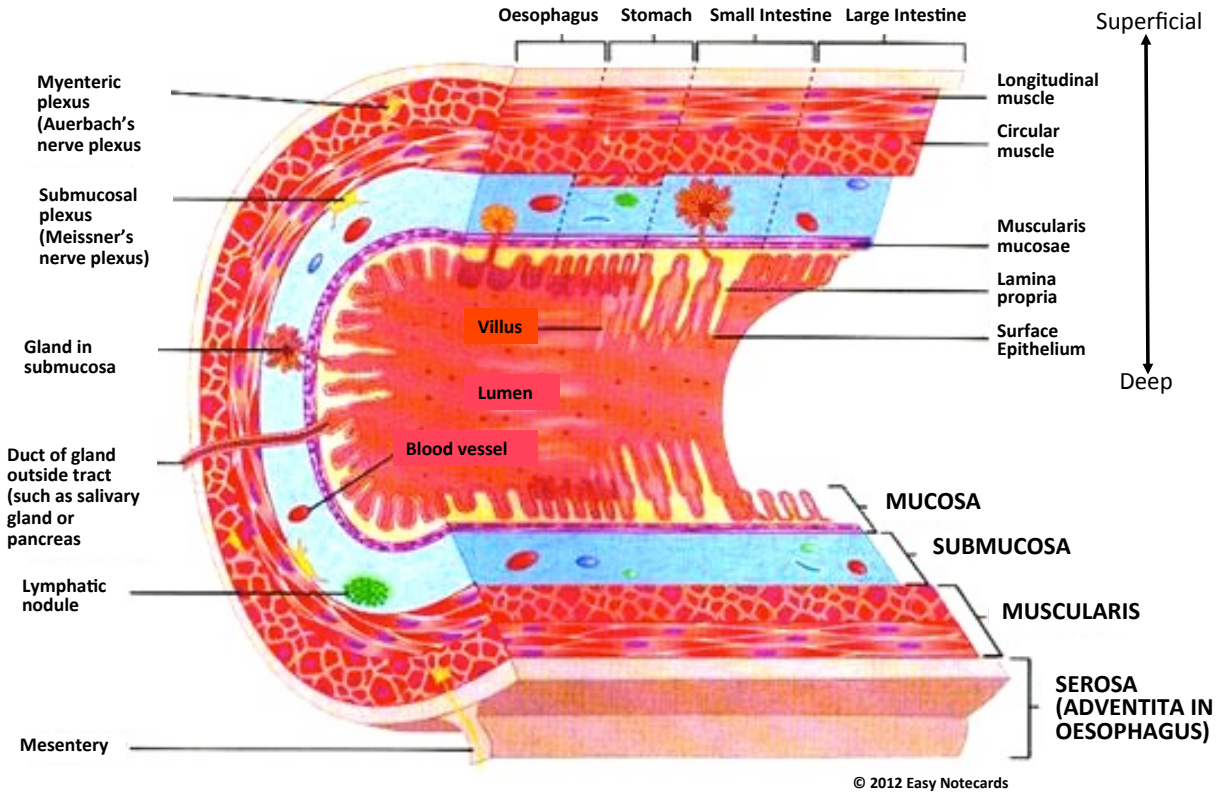


Figure 3.2. Sectional views of layers of the GI tract

The four main layers of the GI tract are depicted, as well as some structural differences in the GI wall in the oesophagus, stomach, small and large intestine.

The *serosa* is the outermost layer of the GI wall. It is made up of a thin layer of loose connective tissue, covered by a type of squamous epithelium referred to as the mesothelium. In the oesophagus, it is termed the *adventitia*. Serous fluid, produced by the serosa lubricates the lining of the GI wall.

The muscular coat that lies between the inner submucosa and the outer adventititia or serosa is the *muscularis*. Along the GI tract, the muscularis usually

consists of an outer longitudinal layer of smooth muscle and an inner circular layer. However, the stomach muscularis contains an additional middle oblique layer. Networks of unmyelinated nerve fibers, called the myenteric plexus, are situated in the muscularis. They control peristalsis, mixing, relaxation of the pyloric and ileocecal spinchters, and smooth muscle tone.

Situated beneath the mucosa is the *submucosa*. It comprises a layer of loose to dense connective tissue containing blood and lymphatic vessels. Another important component of the submucosa is the submucosal plexus, which provides nervous control to the mucosa.

Lining the lumen is the *mucosa*. It differs in structure and function along the length of the GI tract. Epithelial cells cover the mucosa and are in direct contact with the lumen. Different specialized cells dominate the epithelium of the oesophagus, stomach, small and large intestine respectively, contributing to the specialized function of each of these organs. A layer of loose connective tissue (the *lamina propria*) exists beneath the epithelium, and consists of a network of blood vessels and lyphatic nodules beneficial to the immune roles of the GI tract. Dynamic movement of the mucosa is achieved with the *lamina muscularis mucosae*, a thin layer of smooth muscle found after the lamina propria.

Each region of the GI tract has its characteristic histological features (Figure 3.3). Some of the main distinguishing features would be described next.

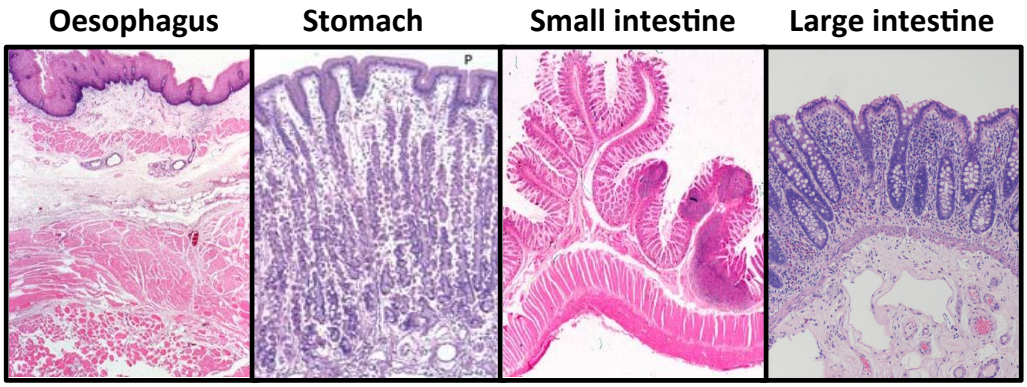


Figure 3.3. Microscopic view of sections of the gastrointestinal tract

Distinctive differences in architecture of the surface epithelium (top region) in the oesophagus, stomach, small and large intestine are depicted. While stratified squamous cells form the epithelial surface of the oesophagus, columnar cells form gastric pits (P) in the stomach. Villi dominate the epithelial surface of the small intestine but is absent in the large intestine which has mucosal folds formed by muscle contractions. Adapted from (8).

The ***oesophagus*** has a *stratified squamous epithelial lining* that limits abrasion during swallowing (Figure 3.3), and is rapidly turned over. In the oesophagus, the *lamina propria* is sparse. And mucous secreted from mucous secreting glands (*papillae*) in the *submucosa* aid the passage of food. The *muscularis propria* consists of layers of smooth and striated muscles that surround the lumen of the oesophagus. These muscle layers control the movement of food down the oesophagus to the stomach by peristalsis.

The ***stomach*** epithelium form deep-folded pits, and contains different types of cells, especially goblet cells. The epithelium is structured to act as a protective lining, to limit autodigestion, and to secrete compounds to facilitate digestion. Unlike other GI organs, the muscularis of the stomach comprises three muscle layers rather than two. These are the inner oblique muscle, circular muscle, and outer longitudinal muscle.

The lining of the ***small intestine*** contains numerous permanent folds called *plicae circulares* and goblet cells. The epithelial surface of each plica is further folded to form numerous *villi*, and the surface of each villus is covered by epithelium with projecting *microvilli*. These maximise the total surface area available for the absorption of water, electrolytes, and nutrients. Large numbers of capillaries present in the villi aid the transfer of digestion products to the hepatic portal vein and the liver. Vessels in the *submucosa* control blood supply to the villi.

Unlike the small intestine, the mucosa of the ***large intestine*** lacks villi. It is specifically structured to reduce abrasion during defecation, and invasion by intestinal microorganisms. The epithelial surface of the colon consists of a single sheet of columnar epithelial cells that form finger-like invaginations at regular intervals, called crypts (See Figure 3.4). Crypts have essential roles in maintaining homeostatic conditions in the large intestine. Alterations in crypt architecture and biology are the underlying causes of most diseases associated with the large bowel.

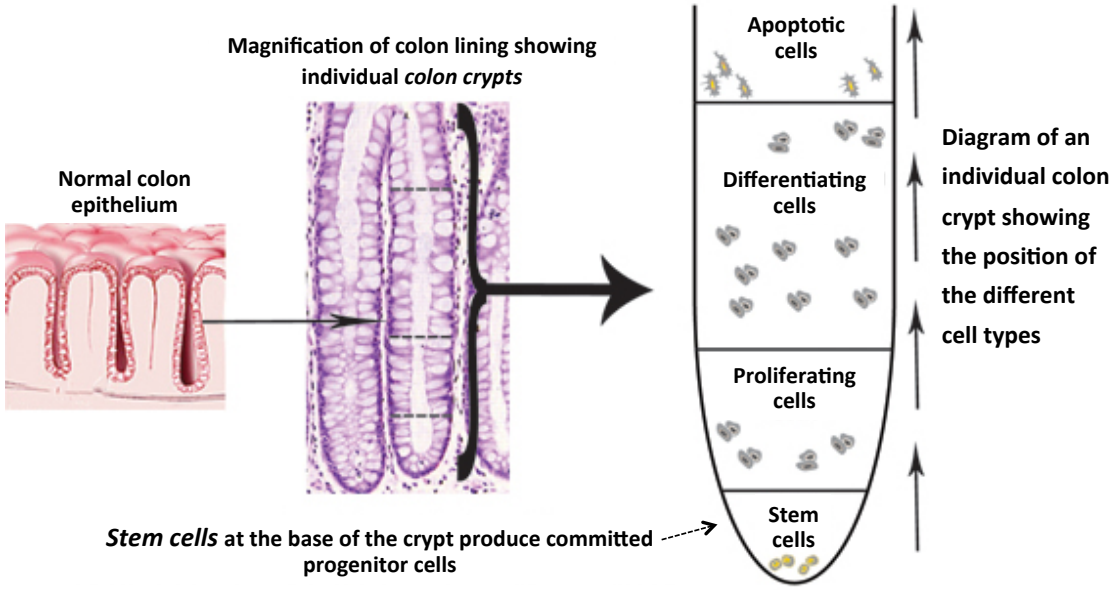


Figure 3.4. Intestinal crypt(s) of the normal colon mucosa.

The normal colon epithelium consists of vertically, closely aligned crypts that extend down the muscularis mucosae at regular intervals. Crypt architecture is displayed in the histological stained section. A representative view of one crypt highlights the location of stem cells, proliferating cells, differentiating cells, and apoptotic cells along the crypt. Adapted from (9).

Intestinal crypts occupy most of the volume of the mucosa, extending vertically from immediately above the muscularis mucosae to open on the mucosal surface. Each crypt contains a range of 500 to 200 epithelial cells, depending on its location in the colorectum (10-13). Mitotically active *proliferating cells* confined to the lower portion of the crypts are stimulated by growth factors involving the Wnt pathway. Epithelial stem cells lie in the base of the proliferative compartment. Committed progenitor cells undergo limited proliferation. As they migrate upward in an ordered fashion, these cells change into *differentiated cells* and replace the older, dying, *apoptotic cells*. Apoptotic cells are shed into the lumen, and the colon epithelium is kept under continuous renewal (10, 12, 14).

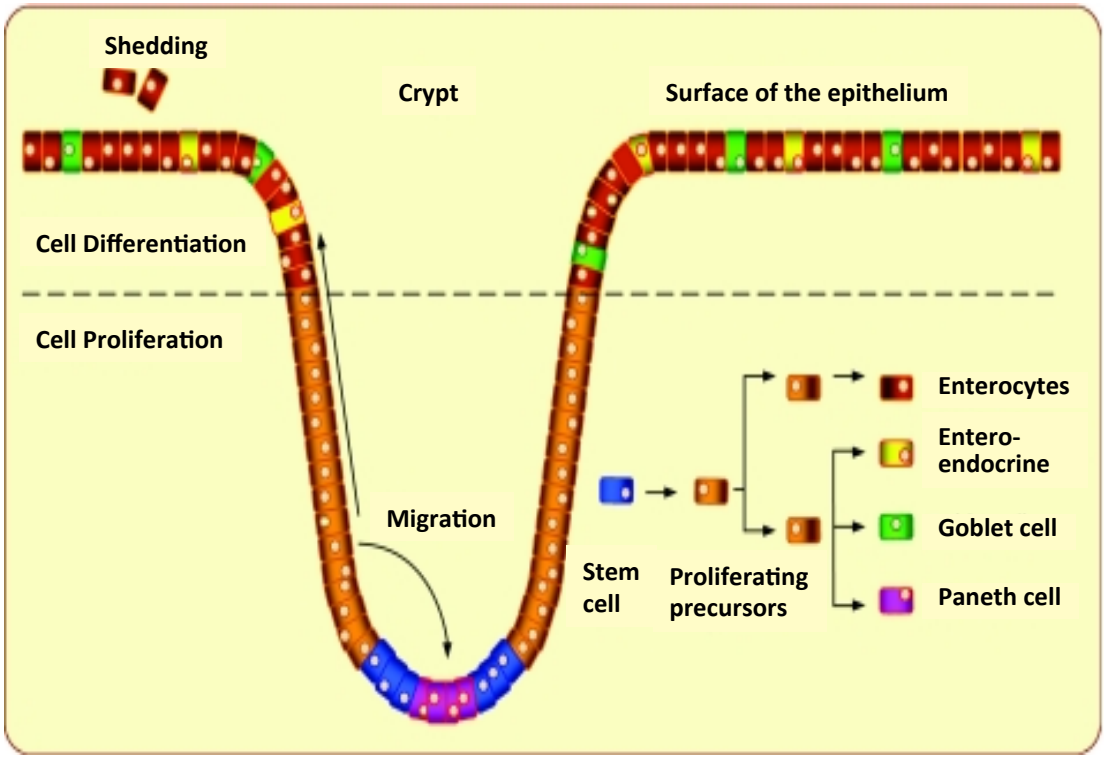


Figure 3.5 Schematic view of the differentiating cells of the colonic epithelium.

Proliferating precursors produced by stem cells migrate towards the crypt top and differentiate into enterocytes (absorptive cells), enteroendocrine cells (hormone secreting cells), goblet cells (mucus secreting cells) or Paneth cells (secrete antimicrobial toxins). Adapted from (14).

As illustrated in Figure 3.5, there are four major types of differentiated cells: the enterocytes or absorptive cells, the mucus-secreting goblet cells, peptide hormone secreting enteroendocrine cells and Paneth cells, which occasionally exist in the ascending colon and in certain disease states (13-15). Paneth cells, secrete antimicrobial toxins, migrate downward and are located at the crypt bottom.

3.3 Diseases of the GI tract (16, 17)

Gastrointestinal diseases constitute a huge burden on health resources worldwide. In the United States, gastrointestinal diseases affect an estimated 60 - 70 million individuals annually (3, 12, 13, 16, 18). Furthermore, a substantial rate of morbidity and mortality from human diseases has been attributed to diseases that affect the organs of the GI tract. The scenario in the United States would most likely be similar in other western countries.

Recent epidemiological reports have identified colorectal cancer (CRC) as the major cause of patient deaths associated with gastrointestinal diseases. Colorectal

cancer accounted for more than half of all GI cancers in the United States and United Kingdom, and is the leading cause of GI-related mortality in most Western countries (3, 5-7). More emphasis would therefore be placed on the biological abnormalities associated with the occurrence of CRC, factors that increase its incidence, and diagnostic tools for the early detection CRC.

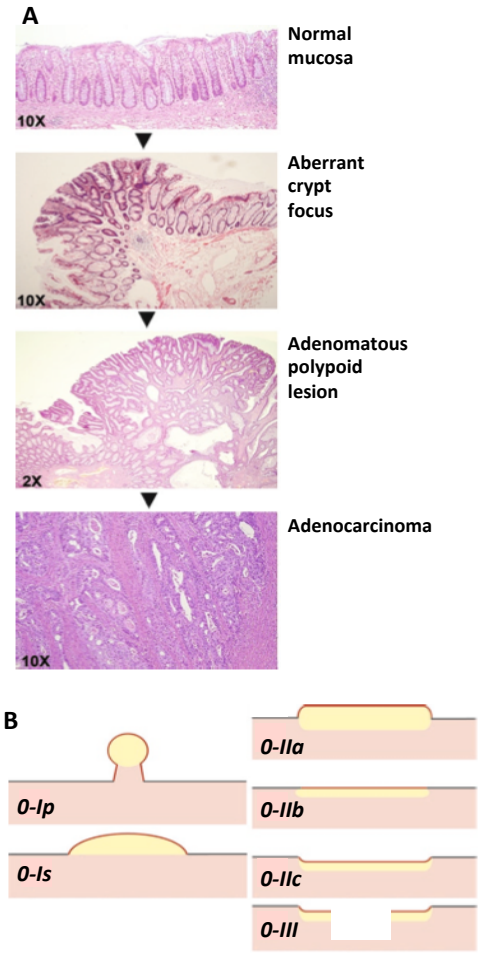
3.3.1 Colorectal Cancer

3.3.1.1 Biology of colorectal cancer

Most sections of the GI tract are prone to precancerous lesions or adenomas. These lesions occur as a result of disruptions in processes regulating normal cell proliferation, cell differentiation, and renewal of the surface epithelium. Precancerous lesions with a characteristic epithelial alteration known as dysplasia are associated with an increased risk of cancer. Dysplasia is a histopathological abnormality characterized by cytological atypia, aberrant differentiation and disorganized architecture (12, 13, 19, 20). Diagnosis and assessment of dysplasia is based mainly on haematoxylin and eosin stained paraffin sections.

In the initial stages of colorectal cancer, the normal epithelium of the large bowel is transformed, forming benign (adenomatuos) lesions (21). Precancerous lesions or adenomas display early morphological changes in discrete clusters of epithelial crypts (12). These neoplastic lesions show some degree of dysplasia, and formation of aberrant crypt foci (ACFs) (10, 12, 13). ACFs are microscopic clusters of few abnormal crypts believed to precede the development of most neoplastic lesions, and are identified by high-magnification-chromoscopic-clonoscopy (15). They are usually larger than normal crypts, have thicker epithelial linings as well as slit-like openings that are raised slightly above the adjacent mucosa (Figure 3.6A) (10, 12). It should be noted that only a small proportion of ACFs progress to neoplastic lesions. Dysplastic ACFs associated with a mutation of the APC gene are associated with the occurrence of ***sporadic colorectal cancer*** through the adenoma-adenocarcinoma pathway (16). Non-dysplastic ACFs occur in the normal colon and are not regarded as precancerous lesions. However, non-dysplastic ACFs with features of hyperplastic polyps and a number of genetic mutations can also lead to colorectal cancer via the less frequented *serrated pathway* (12, 13, 16). Colorectal adenomas resulting from dysplastic ACFs lack invasive capabilities, and most have acquired alterations in the canonical Wnt-APC-βcatenin pathway (13, 19, 20), epigenetic gene silencing, and some degree of

genomic instability (10, 13, 22-24). Over a number of years (5 - 10 years), subsequent accumulation of additional mutations in oncogenes and tumor suppressor genes transforms a minority of these lesions to adenocarcinomas (21). Cattaneo et al., (12) describes a conventional pathway by which this progression is postulated to occur based on consistent scientific findings (Figure 3.6A). In mucosal biopsies stained with methylene blue (Figure 3.6A), ACFs appear darker, larger, and slightly raised above the normal adjacent mucosa. With increasing dysplasia and mutations, a small percentage of ACFs progress to a neoplastic lesion. In sections from the adenocarcinoma (Figure 3.6A), high level of dysplasia, loss of glandular architecture, accumulation of goblet cells and a host of other architectural malformations are noticeable.



D

Type 0 superficial lesions	
Polypoid	
Subtype 0-Ip	protruded, pedunculated
Subtype 0-Is	protruded, sessile
Non-polypoid	
Subtype 0-IIa	Superficial, slightly elevated
Subtype 0-IIb	Flat
Subtype 0-IIc	Superficial shallow, depressed
Subtype 0-III	Excavated (ulcer)

Figure 3.6. Pathway for the transformation of dysplastic adenomas to adenocarcinomas and schematic representation of *type 0* neoplastic lesions.

Histological features of the four stages of the adenoma-adenocarcinoma pathway to colorectal cancer - the normal mucosa, aberrant crypt foci (larger crypts in central region), a tubulovillous adenoma with epithelial dysplasia, and an invasive adenocarcinoma (A). Schematic representation of the main forms of

type 0 neoplastic superficial lesions defined according to the Paris classification (B). The endoscopic appearances of polypoid (pedunculated) and nonpolypoid (slighlty elevated) are shown in (C) and the distinctive characteristics of *type* 0 lesions are described in (D). Adapted from (12, 25).

Colorectal adenoma variants (*adenomas* are hereafter used to describe neoplastic precancerous lesions) are grouped based on endoscopic appearance according to a defined terminology - the Paris endoscopic classification (25), (Figure 3.6B). There are two main groups of the *type 0* neoplastic tumors: polypoid and non-polypoid lesions (Figure 3.6 B, C and D). Polypoid lesions protrude above the surrounding mucosal surface and could be either pedunculated, with a narrow base (*Ip*) or sessile (*Is*), wherein the base and the top of the lesions have a similar diameter (25). The non-polypoid or non-protruding neoplastic lesions are usually either slightly elevated (*IIa*), flat (*IIb*) or depressed (*IIc*) (25). Other types of growth patterns exist. Standard colonoscopy is usually performed to detect and resect these lesions, as they are believed to sometimes progress to invasive adenocarcinomas. Superficial invasive lesions (cancers) are commonly grouped according to the Tumor-Node-Metastasis (TNM) classification (26) (discussed later).

Advanced colorectal adenomas with greater chances of malignancy are often identified based on size (≥ 1 cm), degree of dysplasia, and/or villous histology (16, 27). Distinguishing villous structures from elongated separated tubules could be challenging. Tubular adenomas (TA) usually contain <20% of villous change. Lesions with 20% - 80% villosity have both villous and tubular structures and are referred to as tubulovillous adenomas (TVA). While villous adenomas (VA) lack the tubular growth pattern. Adenomas are more often diagnosed with the TVA pattern. However, the presence of villous changes in adenomas has been linked to increased malignant potential of these tumors, specifically called sessile adenomas (27, 28).

Colorectal cancers could be either of a *sporadic* form (described above) or ***familial / inherited* form** (29). Inherited large-bowel cancers occur in patients with a family history of the disease and are of two major groups: the *polyposis syndromes* and *nonpolyposis syndromes*.

Polyposis syndromes (28-33): This includes familial adenomatuos polyposis, Peutz-Jeghers syndrome and juvenile polyposis. Familial adenomatous polyposis (FAP) is the most frequent, and it accounts for less than 1% of the total CRC burden. It is a rare condition and patients are plagued with numerous amounts of adenomas in the large bowel. Tumors in FAP, are frequent in the distal colon and rectum, more aggressive, and are more likely to have aneuploid DNA. FAP is transmitted though an

autosomal dominant trait. It is associated with a deletion in the long arm of chromosome 5 (which contains the adenomatuous polyposis gene, *APC*), in both neoplastic and normal cells. Allelic loss of this genetic material and mutations in *p53*, and *KRAS* lead to a lack of tumor-suppressor genes and proteins that inhibit tumor growth. Upon confirmation of the disease with molecular genetic testing, patients are recommended for routine CRC screening between ages 20 - 55. FAP defects could also result in Gardner's syndrome and Turcot's syndrome.

Nonpolyposis syndrome (28-31, 34, 35): Hereditary nonpolyposis colon cancer (HNPCC), also referred to as *Lynch syndrome*, is an autosomal dominant trait, characterized by a pertinent family history of CRC. About 1-6% of all colorectal cancers result from HNPCC. For HNPCC patients, a pattern of primary cancers (CRC inclusive) would have been clinically diagnosed before the age of 50 in three or more relatives spanning at least two generations. Also, at least one of the affected family members would be a first-degree relative of the other two. Germ-line mutations of various genes occur in HNPCC. The most predominant are the mismatch repair genes, in particular *hMLH1* on chromosome 3, *hMSH2* on chromosome 2, as well as *hPMS1* and *hPMS2*. These mutations lead to microsatellite instability, greatly comprise cellular DNA replication and DNA mismatch repair mechanisms, and leads to errors that promote tumor cell growth, proliferation and differentiation. Unlike FAP, HNPCC patients present a high frequency of less-aggressive cancers arising in the proximal colon. Furthermore, these patients have adenocarcinomas early in life compared to those with sporadic cancers, and usually undergo routine colonoscopy beginning at 25 years.

3.3.2 Colorectal cancer incidence and staging

According to global statistics on cancer incidence and mortality, colorectal cancer still impacts a huge problem on humans and health care systems (36-41). Worldwide, It is the third most commonly diagnosed cancer in males and the second in females (38). While about 80% of the total number of CRC cases are sporadic, the remaining 20% are familial or inherited forms (29). A worrisome fact on the incidence of sporadic CRC is its increase in many historically low risk areas such as Eastern Asia (38) and Africa (42, 43). Although global CRC death rates remain among the top four cancers, it has decreased in recent years in some Western Countries due to increased awareness, early detection and improved treatment (38). The situation is unfortunately different in many countries with limited resources and health infrastructure, where CRC incidence continues to increase. As observed with most cancers, the survival rate for CRC is

highest in individuals diagnosed with the early stage of the disease (37, 44) (Figure 3.7). Irrespective of race, the 5-year survival rate for localized CRC at diagnosis is greater than 90% (36, 37). With regional diagnosis, this drops to 63% (36, 37). Patients diagnosed with distant metastasis have the worst prognosis, with a 5-year survival rate of less than 10%. CRCs spread to the regional lymph nodes and other organs through the portal venous circulation. In most CRC cases, the liver is often the initial organ for distant metastasis.

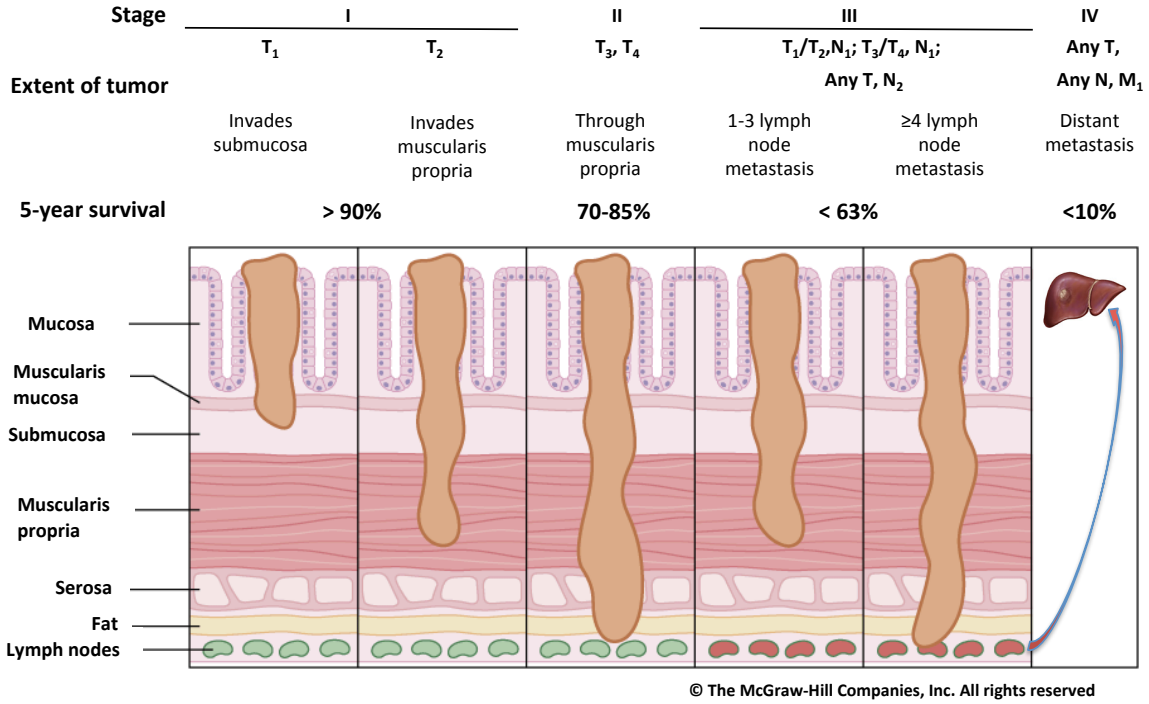


Figure 3.7 TNM stages for colorectal cancer and patient prognosis.

The grading of colorectal cancer based on the TNM system is depicted along with patient prognosis at each stage, and a visual description of stages based on the degree of tumor penetration of the epithelium, lymph node and liver metastasis. Adapted from (28).

The staging of CRC is important as a benchmark for describing CRC occurence, and is useful for clinical, pathological and research purposes. CRC staging is most commonly done using a global standard staging system, known as the TNM system (26). The TNM classification of carcinomas of the colon and rectum is based on the extent of the tumor (T), the degree of spread to the lymph nodes (N), and the presence of metastasis (M). This is illustrated in (Figure 3.7B).

A

T	Presence of tumor, degree of invasiveness
T0	No evidence of primary tumor
Tis	Tumor *in situ* (no extension through the muscularis mucosae)
T1	Tumor invades submucosa
T2	Tumor invades muscularis propria
T3	Tumor invades through the muscularis propria into the subserosa or perirectal tissues
T4	Tumor directly invades other organs and/or perforates visceral peritoneum
N	**Degree of spread to the lymphnodes**
N0	No regional lymph node metastasis
N1	Metastasis in 1 to 3 regional lymph nodes
N2	Metastasis in 4 or more regional lymph nodes
M	**Presence of metastasis**
M0	No distant metastasis
M1	Distant metastasis

B

Stage	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
	T2	N0	M0
IIA	T3	N0	M0
IIB	T4	N0	M0
IIIA	T1-T2	N1	M0
IIIB	T3-T4	N1	M0
IIIC	Any T	N2	M0
IV	Any T	Any N	M1

Figure 3.8 TNM classification of colorectal cancer stages.

Description of the TNM grading system for colorectal cancer (A) and colorectal cancer stages classified according to the TNM scheme (B). Adapted from (25).

Depending on the TNM classification of colorectal cancer, an overall stage of 0, I, II, III, IV can be assigned to an adenocarcinoma (Figure 3.8). These stages could also be subdivided using letters like IIA and IIB. Stage I colorectal cancers are the least advanced, and very often have a better prognosis compared to the more advanced stages (Figure 3.7 and Figure 3.8).

3.3.3 Risk factors for colorectal cancer

A variety of factors contribute to the risk of colorectal adenomas progressing to malignant tumors. This includes lifestyle, diet type, age, exposure to certain diseases, and family history.

Healthy living and adequate physical activity promote fitness of body and mind. The majority of evidence indicates that CRC is a tobacco-associated disease (45-48). High body mass index (BMI) and obesity also increase the risk of colorectal cancer. Studies have also linked high physical activity to reduced risks for CRC (49-55).

Certain dietary habits have also been reported to have an inverse relationship with CRC risk. For example, high alcohol intake predisposes individuals to CRC (56). More so, populations consuming food high in meat and fat, but low in starch, fibre, fruit and vegetables, show higher rates of cancer of the colon and rectum (57-59). The intake of certain chemopreventive agents such as aspirin, calcium, vitamin D and folate have been suggested to avert adenoma formation (45). However more conclusive studies are required to substantiate these claims.

The diagnosis of sporadic CRC is unusual before the age of 40, although its incidence increases exponentially with age (44), with majority of the cases occurring in older individuals over 60 years (36). This is not the case in individuals highly predisposed to CRC due to family history, genetic disposition or certain ailments. For this group of people, CRC risk could be elevated in the late teenage years.

Patients with persistent inflammatory bowel diseases (ulcerative colitis or Crohns' disease) or a long family history of CRC have a higher predisposition to CRC. Furthermore, individuals with some types of inherited gene defects are at risk of developing CRC at a young age. For example, persons with familial adenomatous polyposis (FAP) syndrome have a germline mutation of the major Wnt signalling pathway component, APC (60). As a result, they develop numerous adenomas at an early age and a few of these lesions may progress to cancer (61). An inherited defect in the DNA mismatch repair genes (*hMLH1*, *hMSH2*, *hPMS1*, *hPMS2*) causes Lynch syndrome, or Hereditary non-polyposis colon cancer (HNPCC) (62). The lifetime risk for CRC in such individuals could be as high as 80%. A positive association has also been demonstrated between the occurrence of Type II diabetes and CRC, due to an overlap of risk factors pertaining to age, physical activity and body mass index (BMI) (44).

3.3.4 Colorectal cancer screening

Health guidelines on CRC recommend screening for individuals, beginning at the age of 50. Of course, high-risk persons should be screened earlier and more often. The reduction in CRC incidence reported in a number of Western countries has been attributed largely to the early detection of CRC: that is at the adenoma or early carcinoma stage (36, 37, 63). For example, in the UK, a one-time flexible sigmoidoscopy screening between 55 and 64 years of age reduced CRC incidence by 33% and mortality by 43% (64). The main screening techniques employed in CRC detection would be further described (45, 63, 65-67).

Screening tests that can detect both colorectal adenomas and cancers are frequently recommended to patients. This category of tests is described below.

Flexible sigmoidoscopy involves an examination of the rectum and **parts** of the colon (the distal colon) with the aid of a sigmoidoscope. Tumors detected in the distal colon and rectum examined can be removed.

With *colonoscopy*, the **entire** length of the colon and rectum can be visually examined. This more detailed procedure improves the detection and removal of adenomas in the proximal colon.

The *double-contrast barium enema* (DCBE) test is a type of x-ray test performed with a chalky liquid called barium sulfate. X-ray pictures of the lining of the colon are taken to search for adenomas or cancers. A positive DCBE is usually followed up with a colonoscopy test for resection of the lesion.

CT colonography or virtual colonoscopy is an advanced type of computed tomography scan of the colon and rectum. This technique is more sensitive than DCBE, as computerized images with both two- and three dimensional- views of the inside of the colorectum can be observed for neoplastic lesions. If positive, a colonoscopy would most probably be needed to remove the tumor.

One **major pitfall** with the tests described so far is that they are invasive, expensive, time-consuming, and individuals are exposed to small amounts of radiation. A more worrisome fact is that studies consistently report that patient compliance is low, and almost 70% of individuals who undergo colonoscopy have no colorectal neoplasia (45, 68-70).

This highlights the need for a better diagnostic approach to estimate the absolute risk for individuals who undergo these screening techniques, and to more efficiently target those with late adenomas or early adenocarcinomas.

The other types of screening tests available to the population detect mainly cancerous lesions. These tests involve assessing faeces (stool) samples for signs of cancer.

Fecal occult blood test (FOBT) detects blood in faeces using a guaiac test for the peroxidase activity of haem in haemoglobin. It is the most commonly used CRC screening test. The blood is assumed to be from blood vessels on the surface of large tumors damaged during defaecation. FOBT lacks adequate specificity and sensitivity to detect early, small-sized colorectal tumors. The false positive rate is high as faecal blood could also be from other causes such as ulcers, IBDs, or certain dietary

components. Moreover, not all tumors bleed. When positive, further tests such as colonoscopy have to be performed.

Immunochemical faecal occult blood test (iFOBT) or faecal immunochemical test (FIT) is used to screen faecal samples for a part of the human haemoglobin protein that is found in red blood cells. It uses antibodies against the globin component of human haemoglobin. Like FOBT, FIT may produce false-positive results (albeit less than FOBT) and needs to be followed up with a colonoscopy.

Lastly, *Stool DNA tests* can detect certain aberrant genetic markers such as DNA, released from tumor cells (71-73). A possible drawback for this test is the degradation of long DNA (a marker for non-apoptotic cellss) shed from neoplastic cells by faecal bacterial endonucleases. Despite the improved sensitivity and sensitivity recently reported for this test (68), its performance is generally graded as equal to or lower than the less expensive FIT, and inferior to colonoscopy (74). All three tests are **less invasive**, and **easier to** perform compared to the more invasive CRC screening tests. However, **early stage tumors are frequently missed** (68).

This again emphasizes the necessity for a more sensitive diagnostic method to detect early lesions in colorectal tumorigenesis.

3.4 Somatic alterations in precancerous lesions of the adenoma-carcinoma pathway.

To gain insight on molecular alterations that promote the development of CRC, the biology of adenomatous as well as cancerous lesions must be scrutinized. Numerous gene and mRNA expression studies on these lesions (adenomas and adenocarcinomas) have provided knowledge of molecular alterations that play valuable roles in the development of progressive tumors from normal epithelial cells. In addition to this, non-genetic mutations, especially in DNA methylation have also been discovered from epigenetic studies. Research on these lesions is also driven by the need for biomarkers to aid early detection, diagnosis and monitoring of colorectal tumorigenesis. Tumor-marker-driven studies have targeted extracts from human colorectal tissues, blood and faeces.

3.4.1 Molecular alterations in the genome of adenomas

Contributions to genomic instability in tumor cells are mainly from microsatellite instability (MIN) or chromosomal instability (CIN).

MIN is common in MMR deficient neoplastic lesions from individuals with HNPCC. MIN is observed in about 10-15% of sporadic cancers with acquired transcriptional silencing of *hMLH1* due to the hypermethylation of its promoter region (75-78).

A number of chromosomal abnormalities (CIN) occur in the early stages of colorectal tumorigenesis. Stages involved in the transformation of benign hyperplastic lesions to advanced adenomas and then adenocarcinomas are depicted in Figure 3.9. The loss of chromosome 17p is infrequent in adenomas of any stage (5 - 30%) whereas more than 75% of colorectal carcinomas contain a loss in a large portion of chromosome 17p (79). More importantly, patients with the 17p allelic losses were associated with a high risk of tumor progression from a benign to a malignant state (33). Interestingly, the common region of loss on chromosome 17p in colorectal tumors has been identified to contain *p53*, a tumor suppressor gene frequently switched off in colorectal cancers (21, 80). Another region of allelic loss in colorectal tumors, chromosome 18q, is found in almost 50% of advanced adenomas, and 73% of adenocarcinomas (32). This region has also been mapped, and the candidate tumor suppressor gene identified as *DCC* (21). It encodes a cell adhesion protein believed to be involved in maintaining normal cell adhesion matrix. Gains in chromosomes 8q, 13q and 20q, as well as losses in chromosomes 8p and 15q are also strongly associated with advanced adenomas (81).

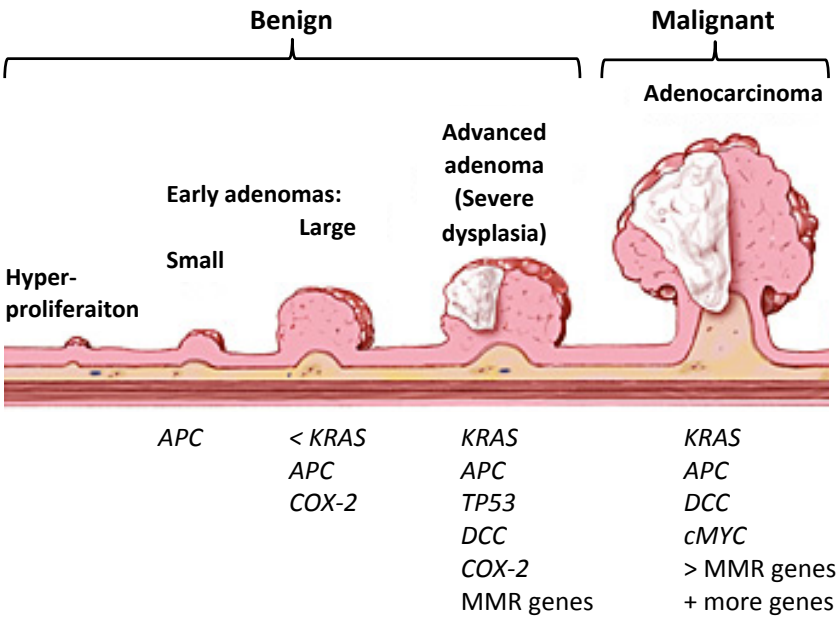


Figure 3.9 Mutations in sporadic colorectal tumorigenesis

Stages involved in the transformation of hyperplastic polyps to early adenomas, advanced adenomas (preinvasive, dysplastic lesions), and adenocarcinomas (invasive lesions). Adapted from (82).

3.4.2 Molecular alterations in the transcriptome of adenomas

Remarkable differences identified in the transcriptome of adenomas when compared to the normal mucosa have revealed evidence of broad-scale transcriptome remodelling in precancerous neoplasms.

The most common gene expression change characteristic of early adenomas is mutations in the *APC* tumor suppressor gene leading to activation in the Wnt signalling pathway. This mutation promotes chromosomal instability and is prevalent in almost 80% of adenomas and 93% of carcinomas (83). The APC tumor suppressor protein regulates intracellular β-catenin levels by facilitating its ubiquitination and subsequent degradation. In the absence of APC, β-catenin is abundant in the cytoplasm, increased amounts are abnormally translocated to the nucleus, where it drives the transcription of various genes involved in tumor growth and invasion (83).

Activating mutation in *KRAS* is another key modification identified in tumor cells. About 50% of colorectal adenomas greater then 1 cm in size were found to have *KRAS* gene mutations (32). While only 9% of sporadic adenomas less than 1 cm showed *KRAS* mutations, approximately 50% of colorectal carcinomas were also found to display this mutation (32, 79, 84, 85). *KRAS* gene mutations could be a necessary feature for small adenomas to progress to an advanced stage, either as an initiating event or through the clonal expansion of mutated cells (79). *KRAS* mutation is predominant in DNA mismatch repair (MMR) proficient tumors and is associated with the promoter methylation of *MGMT* (86). Activating mutations in *EGFR*, and *BRAF* also occur in some adenomas and are suggested to aid a constitutively active RAS signalling pathway.

COX-2, the inducible isoform of *COX*, is expressed in response to mitogens, growth factors and inflammatory stimuli (87). Increased *COX-2* expression has been consistently associated with early tumorigenesis (88, 89). While *COX-2* mRNA levels were detected at high levels in adenoma and adenocarcinomas than in the normal mucosa, no overexpression was found in hyperplastic polyps (89).

In one study, significant expression changes in 78 genes involved in the Wnt signalling cascade was confirmed (90). Also, noteworthy candidates in the list of overexpressed mRNAs in adenomas include four members of the REG family of genes (involved in tissue mitogen expression), *CDH1* (important in epithelial-mesenchymal transition), *CDH3* (associated with early transformation events), *RTEL1* (facilitates telomere elongation) and *KIA1199* (a novel Wnt target).

Additionally, tumor-associated genetic pathways altered at the different stages of adenoma-carcinoma transformation have been identified in comprehensive transcriptomic studies on colorectal adenomas (91). Pathways significantly altered in early colorectal lesions in one of such study are depicted in Figure 3.10.

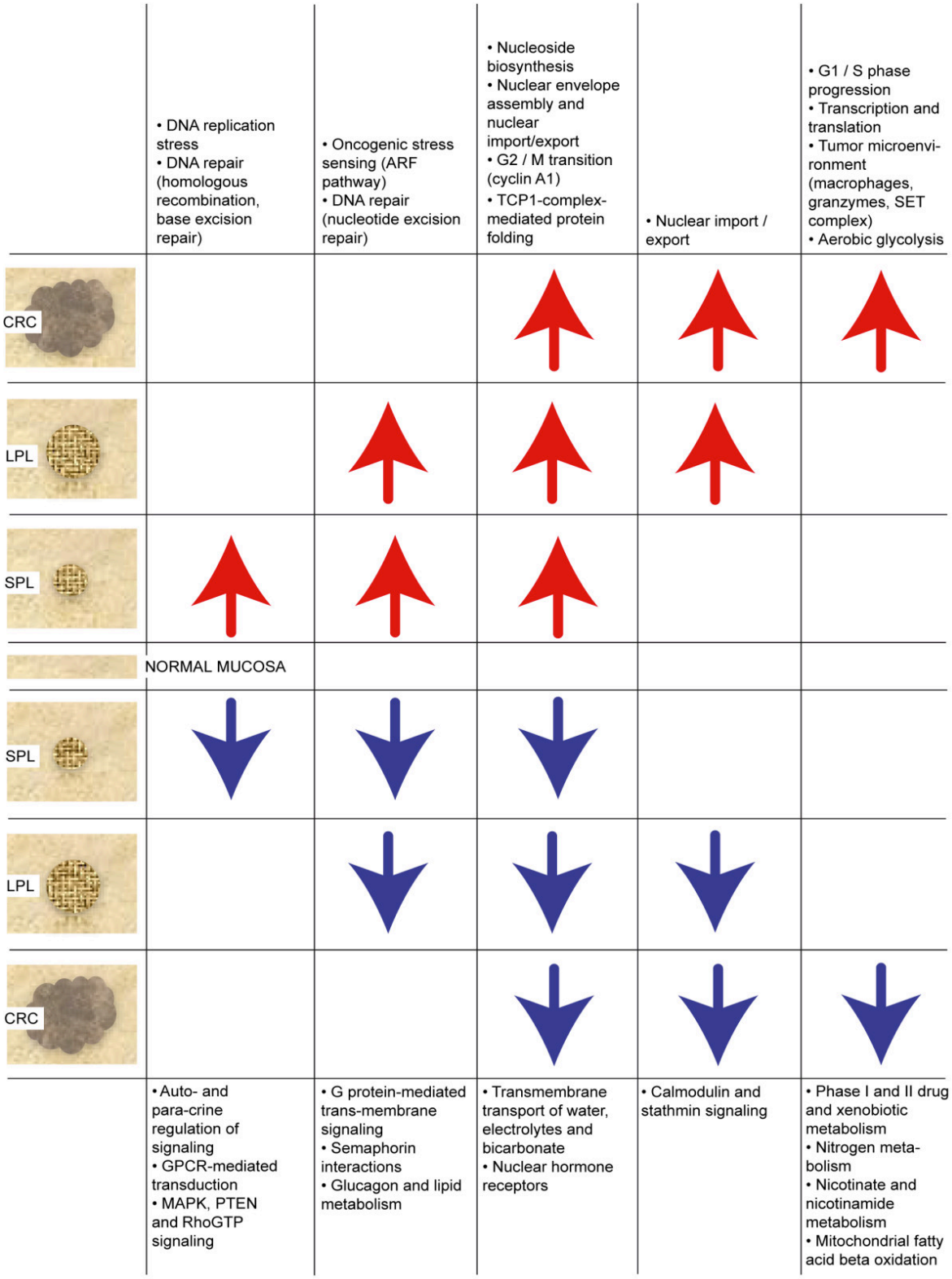


Figure 3.10 Molecular pathways dysregulated during adenoma transformation

An overview of tumor-related pathways identified from transciptomic studies to be dysregulated at different stages (SPL = early adenomas, LPL = advanced adenomas) of colorectal tumorigenesis. Arrows indicate upregulation (up) and downregulation (down). Adapted from (91).

The expression level of transcription factor genes also change during tumorigenesis (92). For example, unlike carcinomas and normal mucosa, adenomas overexpress *DACH1*, which controls the transcription of several genes involved in cell proliferation and differentiation. Downregulation of *TGFB1* in adenomas could also reflect a decline in the function of this proapoptotic gene, which maintains homeostasis in the normal epithelium. Furthermore, the transcription factor genes *BIRC5*, *NR3C1*, and *MYB*, were reported to show significant changes in expression pattern in precancerous lesions.

These gene expression changes represent just a few of the many genetic alterations linked to early precancerous colorectal lesions (93).

3.4.3 Epigenetic alterations in colorectal adenomas

DNA methylation plays crucial roles in the regulation of gene expression. Neoplastic cells have an altered developmental program, and studies have revealed methylation patterns in these cells often deviates from that observed in normal cells.

Goelz et al, showed that total DNA hypomethylation occurs in small, benign colorectal neoplasms, suggesting that an alteration in DNA methylation pathway could be an early event in the initiation of neoplasia (94). Preinvasive lesions also contain more precise hypomethylation patterns. Reduction in genomic 5-methylcytosine observed in adenomas and adenocarcinomas was significantly different from the normal colon mucosa even though the level of hypomethylation was similar in both neoplasms (95). Epigenetic changes such as hypomethylation could also contribute to instability in the cell genome and alter the rate at which genetic alterations occur in tumor cells (79).

Despite global hypomethylation reported for colorectal neoplasms, several genes on the short arm of chromosome 11, including the calcitonin gene are extensively hypermethylated in benign colonic lesions prior to malignant transformation (96). Furthermore, silencing of *SFRP* (the gene encoding secreted frizzled-related proteins) by promoter hypermethylation occurs early in colorectal tumorigenesis and facilitates

constitutive Wnt signalling (97). Increased gene methylation could be of particular importance in the reduction of gene expression.

3.4.4 Metabolic alterations in colorectal adenomas.

Distinctive features in the metabolome of cancer cells were first described in the Warburg effect (98). Today, it is widely accepted that cancer cells undergo a comprehensive metabolome reprograming, influenced by oncogenes, tumor suppressor genes and tumor-associated signalling pathways (93, 99-101). This reprograming is channelled towards increased nutrient uptake and biosynthesis of macromolecules. For example, mutation in *p53* promotes the pentose phosphate pathway and oxidative phosphorylation (99). Likewise, tumor-associated mutations in metabolic enzymes such as isocitrate dehydrogenase and pyruvate kinase 2 isoenzyme are advantageous to cancer cells (100, 101).

Literature on metabolic profile changes in colorectal adenomas is limited. A urine-based diagnostic test comprising butyrate, serine, methanol, β-alanine, p-methylhistidine, 3-hydroxybutyrate, asparagine, trigonelline, 3-hydroxyphenylacetate, histidine, acetone, 2-oxoglutarate, ethanolm adipate, 3-hydroxymandelate, tyrosine and benzoate, identified patients with adenomatous polyps with a greater level of sensitivity (83%) than commercially available faecal-based tests (102). Also, differential separation in the metabolomic profile of adenoma and nonadenoma controls was observed with a panel of 23 metabolites (103). Remarkable findings in this study include an increase in prostaglandin E_2 in adenomas, and a decrease in anti-oxidant-related metabolites 5-oxoproline and diketogulonic acid (103). The overproduction of protaglandin E_2 (PGE_2) is believed to promote the induction of *COX-2* (87, 104). High levels of PGE-M (a major urinary metabolite of protaglandin E_2) has also been linked to increased risk for multiple or advanced adenomas (104).

3.4.5 Proteome alterations in colorectal adenomas

The explanation of an observed phenotype related to a disease should not be based only on genetic and epigenetic changes. Moreover, the environment plays a crucial role in defining phenotype variations. Therefore, proteins associated with these genetic changes need to be unravelled to proffer conclusive relationships between genetic and phenotypic changes.

3.4.5.1 Proteomic studies on colorectal adenomas:

Phenotypic variations in the digestive-tract mucosa, as displayed in endoscopy, occasionally define the different stages associated with colorectal cancer (see Figure 3.7 and Figure 3.9A).

The early diagnosis of colorectal tumors has been greatly facilitated by screening methods based on fecal analysis or colonoscopy, but both approaches have limitations (see Section 3.3.3). Promising results for the detection and validation of potential cancer biomarkers are emerging from proteomic studies on cancer development (105).

In one of the earliest studies on the proteome of colorectal lesions, Stulik et al, (106) analysed protein extracts from normal and adenocarcinoma human tissue specimens using two-dimensional gel electrophoresis (2DE). Subsequent reports introduced a one dimensional electrophoresis step based on iso-electric focusing prior to 2DE or modifications in 2DE, such as two dimensional differential gel electrophoresis (2D-DIGE) (107-112). These reports have primarily evaluated the protein profiles of normal and adenocarcinoma biopsies, with the exclusion of precancerous adenomas. Proteins and pathways differentially regulated in adenoma compared to normal tissues could be conventional biomarker targets for the early detection of colorectal tumors in contrast to those up- or down-regulated in carcinoma/normal tissue. In the first reported study targeting a relatively large number of adenomatuous and matched normal mucosa lesions, the high abundance proteins ANXA3, S100A11, EIF-5A1, and S100P, were significantly overexpressed in benign lesions (113). Despite the use of a 2DE-based approach, useful insights were gained on the proteome of precancerous lesions.

In line with technological advances in the field of proteomics, high-throughput proteomic methods could be applied in colon cancer study. Mass spectrometry (MS) based studies has become the hallmark of most proteomic investigations and advancements in MS techniques has greatly improved the depth of proteome coverage, reproducibility and sensitivity of MS based approaches (105, 114-118). Compared with the gel electrophoresis-based approaches used in the past, MS-based proteomic techniques are more sensitive, robust, and reproducible (119). They also offer broader proteome coverage, which also comprises less abundantly expressed cellular proteins (116, 118, 120, 121).

Thus far, human colorectal adenomas have been the focus of only few MS-based proteomics studies (122, 123). Apart from the study by Lam et al., using a 2-DE

approach (113), no comprehensive MS-based study targeting a relatively large series of precancerous lesions has been reported. This highlights the need for further research applying improved MS-based techniques to investigate precancerous lesions. In addition to improving present knowledge on the biology of this tumor phenotype, putative tumor markers in the early stages of colorectal cancer development could be identified.

3.4.6 Mass spectrometry-based proteomic techniques for biomarker studies (114, 115, 124-126).

As illustrated in Figure 3.11, there are two major forms of MS-based proteomic techniques employed in conventional biomarker-centred studies: *shotgun (discovery) MS* and *targeted MS*. Each method can be further classified based on the desired approach for the study and the type of mass spectrometer involved. More importantly, the type of biological sample and problem(s) to be investigated strongly determines which MS method to implement in a biomarker study.

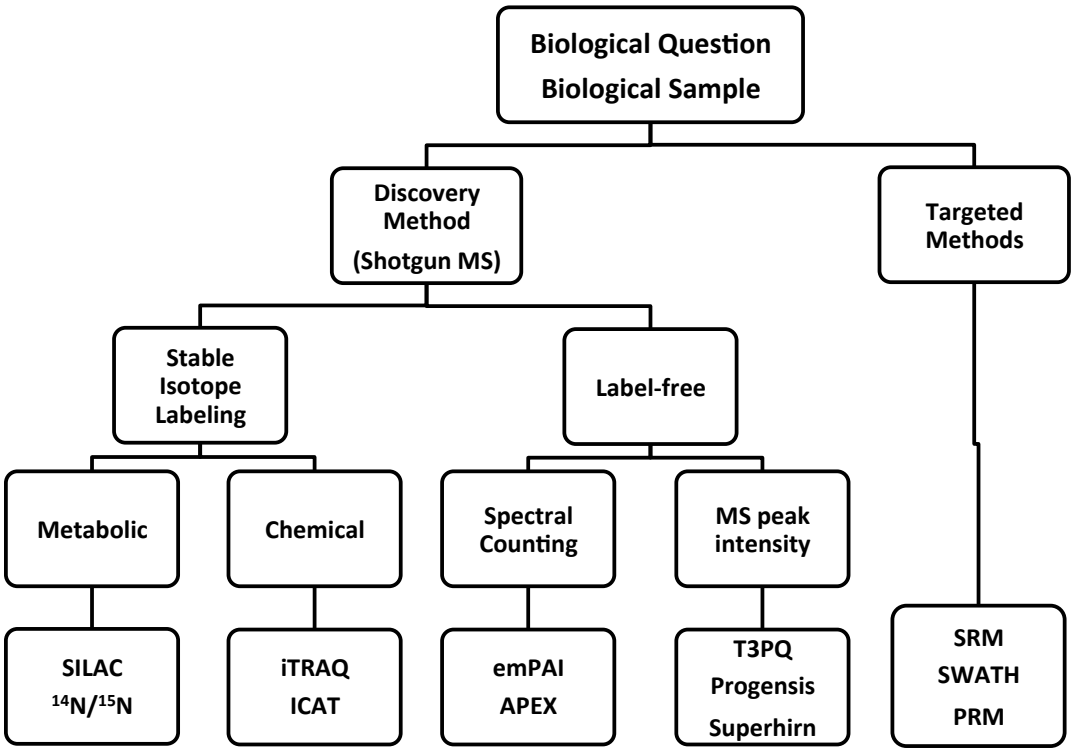


Figure 3.11 Methods applied in mass spectrometry based proteomics

3.4.6.1 Shotgun mass spectrometry (Shotgun MS)

This technique aids global protein identification in the target sample. A typical proteomic workflow for shotgun analysis is shown in Figure 3.12. For a defined

biological research question, the appropriate sample (such as plasma, tissue, or urine) to be analysed is selected.

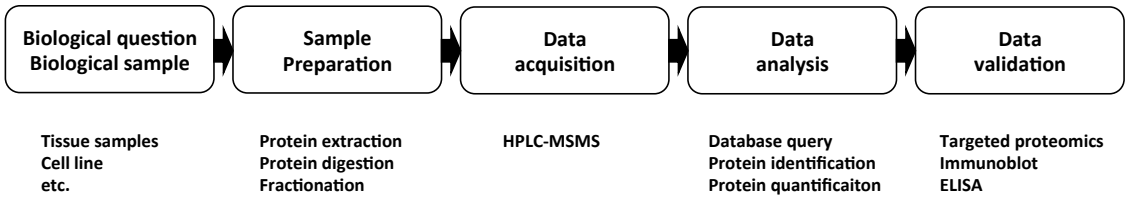


Figure 3.12 A representative proteomics workflow

Sample preparatory steps involve protein extraction, tryptic digestion, and peptide fractionation, to obtain suitable analytes (tryptic peptides) for MS analysis. The peptide fractionation process is crucial for improved proteome coverage as it reduces sample complexity and facilitates the detection of low abundance proteins (127). Coupling a chromatographic apparatus to a mass spectrometer permits direct MS analysis of peptides as they elute from the chromatographic column. Peptides are ionized prior to the first MS phase (MS1) and each charged peptide has a defined mass-to-charge ratio (m/z). In MS1, the m/z and relative abundance of peptides (precursor ions) eluting from the column at defined times are measured and reflected in a mass spectra. Thereafter, the peptide is fragmented and in the second MS phase (MS2), the relative abundance and m/z ratio of fragments ions are measured and documented in the resulting spectra. This type of analysis is called tandem MS or MSMS and results in a number of chromatographic peaks. Peptide spectra features serve as coordinates for peptide detection. Peptide spectra are queried against databases, and matched to the corresponding peptide and proteins, using defined algorithms. The selection of peptide precursors for MS2 is usually restricted to the most intense MS1 peaks (data dependent acquisition, DDA). A variety of mass spectrometers can be used for shotgun proteomics, but data generated from these instruments must be acquired in a data dependent mode. Rather than just identify proteins, quantitative information about all proteins in a sample can be investigated.

To differentiate peptide/protein abundance in different samples, a quantitative dimension could be incorporated (illustrated in Figure 3.11). First of all, with label-free quantification, precursor signal intensity or spectral counting (number of spectra for a given peptide) is used as indicators to measure and compare peptide intensity in different samples. Computational tools like Absolute Protein Expression (APEX), exponentially modified protein abundance index (emPAI), Top 3 Protein Quantification

(T3PQ), Progenesis or SuperHirn could be used to quantify protein abundance in samples (128-131) (Figure 3.11). The second approach involves the use of stable isotope labelling (Figure 3.11). Isotopic labels attached to protein cross-linkers cause a known mass shift in the mass spectrum of labelled peptides. Samples tagged with different labels are usually pooled and analysed. Differences in the peak intensity of isotope labels reflect differences in protein abundance for the samples analysed. Examples of stable isotopes employed for shotgun proteomics are stable isotope labelling by amino acids in cell culture (SILAC), isobaric tag for relative and absolute quantification (iTRAQ), and isotope-coded affinity tags (ICAT) (132-134) (Figure 3.11). SILAC is a metabolic label that can be incorporated into *in vivo* samples. iTRAQ and iCAT are chemical labels better suited to *in vitro* experiments, such as with human samples. The type of labelling preferences applicable to a research study strongly depends on the biological questions to be addressed and the type of biological sample to be investigated.

Drawbacks with the shotgun technique include under-sampling, mainly from the stochastic selection of peptide precursors in MS1 for fragmentation. And secondly by random sampling from dynamic exclusion filtering of precursors for fragmentation in the second MS phase. These factors undermine the sensitivity and reproducibility of shotgun experiments.

3.4.6.2 Targeted mass spectrometry (124, 135-138)

The hallmark of targeted MS is the reproducible detection and quantification of a predetermined set of proteins with high reproducibility and sensitivity, in multiple samples. A similar workflow shown in Figure 3.12 is applicable in targeted MS. However, LC-MSMS is performed in a different type of instrument, and sample fractionation is not recommended as it results in decreased sample throughput and could be a potential source of measurement bias/variability. Unlike discovery MS, data acquisition in targeted MS is performed in a triple quadrupole instrument in a data independent mode (DIA). This means that precursor ions selected for fragmentation, as well as fragment ions for detection are pre-set. No time is spent on the acquisition of unwanted spectra. Selected Reaction Monitoring (SRM) is a targeted technique that has been standardized for high accuracy mass measurements in a large series samples. Furthermore, emerging techniques such as Parallel Reaction Monitoring (PRM) and SWATH have also been applied in targeted proteomics and both have favourable future prospects for targeted proteomics (Figure 3.11).

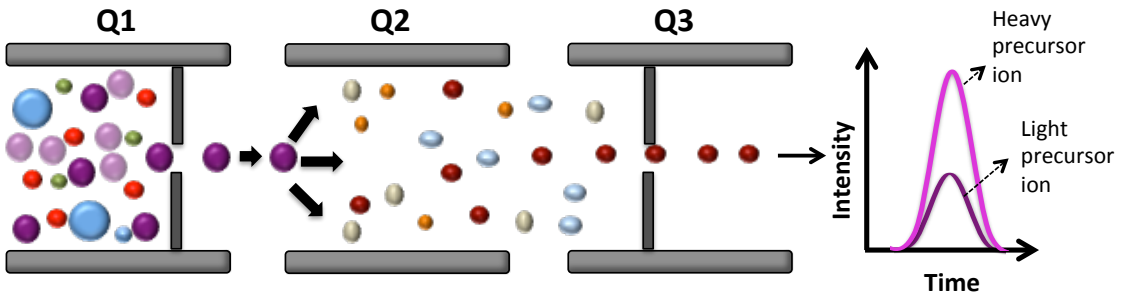


Figure 3.13 SRM analysis in a triple quadrupole instrument

Triple quadrupole (QQQ) instruments are capable of selectively isolating and monitoring specific precursor and corresponding fragment ions - collectively called *SRM transitions*. SRM transitions identify a specific peptide and its corresponding protein. The use of proteotypic peptides, unique to a protein, increases the specificity of SRM measurements. In a typical SRM experiment (see Figure 3.13), a selected precursor ion m/z is detected and isolated in the first quadrupole (Q1). Precursor fragmentation by collision induced dissociation (CID) in the second quadrupole (Q2) creates fragment ions. A set of fragment ions that is selective for the chosen precursor ion is sequentially detected in the third quadrupole (Q3). SRM transitions are monitored over time. The resulting intensity peak and specified elution time (retention time) serve as coordinates to quantify peptide abundance. To obtain an accurate quantification of peptide concentration, defined amounts of synthetic peptide standards (heavy isotope-labelled peptides) could be spiked into samples and monitored along with the analogous endogeneous peptides in the same SRM experiment. At the defined chromatographic elution time, peak intensities for endogenous and labelled peptides are extracted (Figure 3.13) for precise quantitative measurement.

3.4.7 Protein biomarkers for colorectal cancer

At the moment, only two protein-based markers are currently in clinical use for the detection of colorectal cancer. Firstly, faecal occult blood test, which aids the detection of carcinomas albeit with low diagnostic accuracy (discussed in Section 3.3.3). Secondly, carcinoembryonic antigen (CEA) levels are important in evaluating disease prognosis, detecting recurrence, and monitoring therapy in patients with colorectal cancer. Both assays are unsuitable for the detection of adenomas and early stage adenocarcinomas, when disease prognosis is above 90% for patients.

The improved reproducibility, selectivity, sensitivity and dynamic range achieved with high-throughput SRM measurements makes it an ideal technique for the validation of putative biomarker candidates. Literature is vast with potential colorectal cancer biomarkers but very few have progressed to clinical use (93, 123). One reason for this is the harsh terrain that governs the pipeline from biomarker discovery to clinical application. Another major reason is the lack of an appropriate technique for large-scale verification, optimization and validation of biomarker candidates. Presently, the most commonly used validation technique is enzyme linked immunosorbent assay (ELISA). Although it is a quantitative technique, the limitations of ELISA include availability of a suitable antibody, high cost of antibodies, low reproducibility, and small analytical size. If fully optimized, targeted MS methods such as SRM, has the potential to overcome these setbacks. This technique can also be applied for large-scale biomarker verification, and validation in multiple samples.

3.5 References

1. Yamada, T., and Alpers, D. H. (1999) *Textbook of gastroenterology* (Lippincott Williams \& Wilkins)

2. National Digestive Diseases Information Clearinghouse (NDDIC) Gas in the digestive tract (DD-30).

3. Peery, A. F., Dellon, E. S., Lund, J., Crockett, S. D., McGowan, C. E., Bulsiewicz, W. J., Gangarosa, L. M., Thiny, M. T., Stizenberg, K., and Morgan, D. R. (2012) Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology* 143, 1179–1187. e3

4. Talbot, I., Price, A., and Salto-Tellez, M. (2006) *Biopsy Pathology in Colorectal Disease, 2Ed* (Taylor \& Francis)

5. Williams, J. G., Roberts, S. E., Ali, M. F., Cheung, W. Y., Cohen, D. R., Demery, G., Edwards, A., Greer, M., Hellier, M. D., Hutchings, H. A., Ip, B., Longo, M. F., Russell, I. T., Snooks, H. A., and Williams, J. C. Gastroenterology services in the UK. The burden of disease, and the organisation and delivery of services for gastrointestinal and liver disorders: a review of the evidence. *gut.bmj.com*,

6. Russo, M. W., Wei, J. T., Thiny, M. T., Gangarosa, L. M., Brown, A., Ringel, Y., Shaheen, N. J., and Sandler, R. S. (2004) Digestive and liver diseases statistics, 2004. *Gastroenterology* 126, 1448–1453

7. Altobelli, E., Lattanzi, A., Paduano, R., Varassi, G., and di Orio, F. (2014) Colorectal cancer prevention in Europe: Burden of disease and status of screening programs. *Prev Med* 62C, 132–141

8. Anders, R. A. Functional Histology of the Gastrointestinal Tract. *hopkinsmedicine.org*,

9. Colorectal Cancer - Stem Cells & Colon Cancer Colorectal Cancer - Stem Cells & Colon Cancer. *hopkinscoloncancercenter.org*,

10. ALRAWI, S. J., SCHIFF, M., CARROLL, R. E., DAYTON, M., GIBBS, J. F., KULAVLAT, M., TAN, D., BERMAN, K., STOLER, D. L., and ANDERSON, G. R. (2006) Aberrant Crypt Foci. *Anticancer …*,

11. Potten, C. S., Kellett, M., Roberts, S. A., Rew, D. A., and Wilson, G. D. (1992) Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut* 33, 71–78

12. Cattaneo, E., Baudis, M., Buffoli, F., Bianco, M. A., Zorzi, F., and Marra, G. (2010) in *Pre-Invasive Disease: …* (Springer New York, New York, NY), pp 369–394.

13. Humphries, A., and Wright, N. A. (2008) Colonic crypt organization and tumorigenesis. *Nat Rev Cancer* 8, 415–424

14. Schneikert, J., and Behrens, J. (2007) The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut* 56, 417–425

15. Hurlstone, D. P., Karajeh, M., Sanders, D. S., Drew, S. K., and Cross, S. S. (2005) Rectal Aberrant Crypt Foci Identified Using High-Magnification-Chromoscopic Colonoscopy: Biomarkers for Flat and Depressed Neoplasia. *Am J Gastroenterology* 100, 1283–1289

16. Hamilton, S. R., Aaltonen, L. A., World Health Organization, International Agency for Research on Cancer (2000) *Pathology and genetics of tumours of the digestive system* (Oxford University Press, USA)

17. Morson, B. C., Jass, J. R., and Sobin, L. H. (1985) Precancerous lesions of the gastrointestinal tract: a histological classification (Baillière Tindall)

18. NDDIC (2013) Digestive Diseases Statistics for the United States. 1–6

19. Powell, S. M., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., Vogelstein, B., and Kinzler, K. W. (1992) APC mutations occur early during colorectal tumorigenesis. *Nature* 359, 235–237

20. Morin, P. J. (1997) Activation of beta -Catenin-Tcf Signaling in Colon Cancer by Mutations in beta -Catenin or APC. *Science* 275, 1787–1790

21. Fearon, E. R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767

22. Baylin, S. B., and Ohm, J. E. (2006) Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer* 6, 107–116

23. Goss, K. H., and Groden, J. (2000) Biology of the Adenomatous Polyposis Coli Tumor Suppressor. *Journal of Clinical Oncology*,

24. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1997) Genetic instability in colorectal cancers. *Nature* 386, 623–627

25. Anon (2003) The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest. Endosc.* 58, S3–43

26. Greene, F. L., Page, D. L., Fleming, I. D., Balch, C. M., and Fritz, A. G. (2002) *Ajcc Cancer Staging Handbook Plus Eztnm* (Springer)

27. Ishii, T., Notohara, K., Umapathy, A., Mallitt, K.-A., Chikuba, H., Moritani, Y., Tanaka, N., Rosty, C., Matsubara, N., Jass, J., Leggett, B., and Whitehall, V. (2011) Tubular adenomas with minor villous changes show molecular features characteristic of tubulovillous adenomas. *Am. J. Surg. Pathol.* 35, 212–220

28. Longo, D., Fauci, A., Kasper, D., Hauser, S., Jameson, J., and Loscalzo, J. (2011) *Harrison's Principles of Internal Medicine, 18th Edition* (Mcgraw-hill)

29. Lynch, H. T., and la Chapelle, de, A. (2003) Hereditary colorectal cancer. *N. Engl. J. Med*,

30. Fitzsimmons, M. L. (1992) Hereditary colorectal cancers. *Seminars in Oncology Nursing* 8, 252–257

31. Utsunomiya, J. (1990) in *Hereditary colorectal cancer* (Springer Japan, Tokyo), pp 3–16.

32. Vogelstein, B., Fearon, E., and Hamilton, S. (1988) Genetic alterations during colorectal-tumor development. *… England Journal of …*,

33. Fearon, E., Hamilton, and Vogelstein, B. (1987) Clonal analysis of human colorectal tumors. *Science* 238, 193–197

34. Liu, B., Parsons, R., Papadopoulos, N., Nicolaides, N. C., Lynch, H. T., Watson, P., Jass, J. R., Dunlop, M., Wyllie, A., Peltomäki, P., Chapeele, A. D. L., Hamilton, S. R., Vocelstein, B., and Kinzler, K. W. (1996) Analysis of mismatch repair genes in hereditary non–polyposis colorectal cancer patients. *Nat Med* 2, 169–174

35. Lamberti, C., Kruse, R., Ruelfs, C., Caspari, R., Wang, Y., Jungck, M., Mathiak, M., Malayeri, H. R., Friedl, W., Sauerbruch, T., and Propping, P. (1999) Microsatellite instability-a useful diagnostic tool to select patients at high risk for hereditary non-polyposis colorectal cancer: a study in different groups of patients with colorectal cancer. *Gut* 44, 839–843

36. Siegel, R., Naishadham, D., and Jemal, A. (2013) Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians* 63, 11–30

37. Siegel, R., Desantis, C., and Jemal, A. (2014) Colorectal cancer statistics, 2014. *CA: A Cancer Journal for Clinicians* 64, 104–117

38. Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011) Global cancer statistics. *CA: A Cancer Journal for Clinicians* 61, 69–90

39. Chen, W., Zheng, R., Zhang, S., Zhao, P., Zeng, H., Zou, X., and He, J. (2014) Annual report on status of cancer in China, 2010. *Chin. J. Cancer Res.* 26, 48–58

40. Haberland, J., Bertz, J., Wolf, U., Ziese, T., and Kurth, B.-M. (2010) German cancer statistics 2004. *BMC Cancer* 10,

41. Hayat, M. J., Howlader, N., Reichman, M. E., and Edwards, B. K. (2007) Cancer Statistics, Trends, and Multiple Primary Cancer Analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. *The Oncologist* 12, 20–37

42. Jemal, A., Bray, F., Forman, D., O'Brien, M., Ferlay, J., Center, Melissa, and Parkin, D. M. (2012) Cancer burden in Africa and opportunities for prevention. *Cancer* 118, 4372–4384

43. Parkin, D. M., Ferlay, J., Hamdi-Chérif, M., and Sitas, F. (2003) Cancer in Africa. *… and prevention IARC …*,

44. Kim, K. E. (2009) Early Detection and Prevention of Colorectal Cancer (SLACK)

45. Kahi, C. J., Rex, D. K., and Imperiale, T. F. (2008) Screening, Surveillance, and Primary Prevention for Colorectal Cancer: A Review of the Recent Literature. *Gastroenterology* 135, 380–399

46. Chao, A., Thun, M. J., Jacobs, E. J., Henley, S. J., Rodriguez, C., and Calle, E. E. (2000) Cigarette smoking and colorectal cancer mortality in the cancer prevention study II. *J. Natl. Cancer Inst.* 92, 1888–1896

47. Verla-Tebit, E., Lilla, C., Hoffmeister, M., Brenner, H., and Chang-Claude, J. (2006) Cigarette smoking and colorectal cancer risk in Germany: a population-based case-control study. *Int. J. Cancer* 119, 630–635

48. Giovannucci, E. An Updated Review of the Epidemiological Evidence that Cigarette Smoking Increases Risk of Colorectal Cancer. *cebp.aacrjournals.org*,

49. Dai, Z., Xu, Y. C., and Niu, L. (2007) Obesity and colorectal cancer risk: a meta-analysis of cohort studies. *WJG*,

50. Larsson, S. C., and Wolk, A. (2007) Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *The American journal of clinical nutrition*,

51. Vermorken, A. J. M., Zhu, J., and Andres, E. (2014) Obesity and colorectal cancer risk: the role of oxidative stress. *Gut* 63, 529–530

52. Frezza, E. E. (2006) Influence of obesity on the risk of developing colon cancer. *Gut* 55, 285–291

53. Slattery, M. L. (2004) Physical Activity and Colorectal Cancer. *Sports Medicine* 34, 239–252

54. Slattery, M. L. (2003) Physical Activity and Colorectal Cancer. *American Journal of Epidemiology* 158, 214–224

55. Colditz, G. A., Cannuscio, C. C., and Frazier, A. L. (1997) Physical activity and reduced risk of colon cancer: implications for prevention - Springer. *Cancer Causes and Control* 8, 649–667

56. A H Wu, A. P.-H. R. K. R. B. E. H. (1987) Alcohol, physical activity and other risk factors for colorectal cancer: a prospective study. *Br. J. Cancer* 55, 687

57. SA, B. (2000) Diet and colorectal cancer prevention. *Biochem Soc Trans* 28, 12–16

58. Bingham, S. A., Day, N. E., Luben, R., Ferrari, P., Slimani, N., Norat, T., Clavel-Chapelon, F., Kesse, E., Nieters, A., Boeing, H., Tjønneland, A., Overvad, K., Martinez, C., Dorronsoro, M., Gonzalez, C. A., Key, T. J., Trichopoulou, A., Naska, A., Vineis, P., Tumino, R., Krogh, V., Bueno-de-Mesquita, H. B., Peeters, P. H., Berglund, G., Hallmans, G., Lund, E., Skeie, G., Kaaks, R., and Riboli, E. (2003) Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *The Lancet* 361, 1496–1501

59. Terry, P., Giovannucci, E., Michels, K. B., Bergkvist, L., Hansen, H., Holmberg, L., and Wolk, A. (2001) Fruit, Vegetables, Dietary Fiber, and Risk of Colorectal Cancer. *JNCI Journal of the National Cancer Institute* 93, 525–533

60. Sillars-Hardebol, A. H., Carvalho, B., van Engeland, M., Fijneman, R. J. A., and Meijer, G. A. (2012) The adenoma hunt in colorectal cancer screening: defining the target. *J. Pathol.* 226, 1–6

61. Muto, T., Bussey, H. J. R., and Morson, B. C. (1975) The evolution of cancer of the colon and rectum. *Cancer* 36, 2251–2270

62. Chung, D. C., and Rustgi, A. K. (1995) DNA mismatch repair and cancer. *Gastroenterology* 109, 1685–1699

63. Laine, C. (2008) Colorectal Cancer Screening. *Ann Intern Med* 148, ITC2

64. Robert J Hilsden, A. R. (2010) Colorectal cancer screening using flexible sigmoidoscopy: United Kingdom study demonstrates significant incidence and mortality benefit. *Canadian Journal of Gastroenterology* 24, 479

65. Levin, B., Lieberman, D. A., McFarland, B., Smith, R. A., Brooks, D., Andrews, K. S., Dash, C., Giardiello, F. M., Glick, S., Levin, T. R., Pickhardt, P., Rex, D. K., Thorson, A., Winawer, S. J., for the American Cancer Society Colorectal Cancer Advisory Group, the US Multi-Society Task Force, and the American College of Radiology Colon Cancer Committee (2008) Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA: A Cancer Journal for Clinicians* 58, 130–160

66. Pignone, M. (2002) Screening for Colorectal Cancer in Adults at Average Risk: A Summary of the Evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 137, 132–141

67. Davies, R. J., Miller, R., and Coleman, N. (2005) Colorectal cancer screening: prospects for molecular stool analysis. *Nat Rev Cancer* 5, 199–209

68. Imperiale, T. F., Wagner, D. R., Lin, C. Y., Larkin, G. N., Rogge, J. D., and Ransohoff, D. F. (2000) Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *N. Engl. J. Med.* 343, 169–174

69. Betes, M., Munoz-Navas, M. A., Duque, J. M., Angos, R., Macias, E., Subtil, J. C., Herraiz, M., Riva, S., Delgado-Rodriguez, M., and Martinez-Gonza'lez, M. A. (2003) Use of colonoscopy as a primary screening test for colorectal cancer in average risk people. *Am J Gastroenterology* 98, 2648–2654

70. Lieberman, D. A., Weiss, D. G., Bond, J. H., Ahnen, D. J., Garewal, H., Harford, W. V., Provenzale, D., Sontag, S., Schnell, T., Durbin, T. E., Nelson, D. B., Ewing, S. L., Triadafilopoulos, G., Ramirez, F. C., Lee, J. G., Collins, J. F., Fennerty, M. B., Johnston, T. K., Corless, C. L., McQuaid, K. R., Sampliner, R. E., Morales, T. G., Fass, R., Smith, R., Maheshwari, Y., and Chejfec, G. (2000) Use of Colonoscopy to Screen Asymptomatic Adults for Colorectal Cancer. *N. Engl. J. Med.* 343, 162–168

71. Calistri, D., Rengucci, C., Bocchini, R., Saragoni, L., Zoli, W., and Amadori, D. (2003) Fecal multiple molecular tests to detect colorectal cancer in stool. *Clin. Gastroenterol. Hepatol.* 1, 377–383

72. Dong, S. M., Traverso, G., Johnson, C., Geng, L., Favis, R., Boynton, K., Hibi, K., Goodman, S. N., D'Allessio, M., Paty, P., Hamilton, S. R., Sidransky, D., Barany, F., Levin, B., Shuber, A., Kinzler, K. W., Vogelstein, B., and Jen, J. (2001) Detecting colorectal cancer in stool with the use of multiple genetic targets. *J. Natl. Cancer Inst.* 93, 858–865

73. Ahlquist, D. A., Skoletsky, J. E., Boynton, K. A., Harrington, J. J., Mahoney, D. W., Pierceall, W. E., Thibodeau, S. N., and Shuber, A. P. (2000) Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel. *Gastroenterology* 119, 1219–1227

74. Song, K., Fendrick, A. M., and Ladabaum, U. (2004) Fecal DNA testing compared with conventional colorectal cancer screening methods: a decision analysis. *Gastroenterology* 126, 1270–1279

75. Samowitz, W. S., Curtin, K., Ma, K. N., Schaffer, D., Coleman, L. W., Leppert, M., and Slattery, M. L. (2001) Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol. Biomarkers Prev.* 10, 917–923

76. Cunningham, J. M., Christensen, E. R., Tester, D. J., Kim, C.-Y., Roche, P. C., Burgart, L. J., and Thibodeau, S. N. Hypermethylation of the hMLH1 Promoter in Colon Cancer with Microsatellite Instability. *cebp.aacrjournals.org*,

77. Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J.-P. J., Markowitz, S., Willson, J. K. V., Hamilton, S. R., Kinzler, K. W., Kane, M. F., Kolodner, R. D., Vogelstein, B., Kunkel, T. A., and Baylin, S. B. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma.

78. Thibodeau, S. N., French, A. J., Cunningham, J. M., Tester, D., Burgart, L. J., Roche, P. C., McDonnell, S. K., Schaid, D. J., Vockley, C. W., Michels, V. V., Farr, G. H., and O'Connell, M. J. Microsatellite Instability in Colorectal Cancer: Different Mutator Phenotypes and the Principal Involvement of hMLH1. *cebp.aacrjournals.org*,

79. Fearon, E. R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767

80. Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., VanTuinen, P., Ledbetter, D. H., Barker, D. F., Nakamura, Y., White, R., and Vogelstein, B. (1989) Chromosome 17 Deletions and p53 Gene Mutations in Colorectal Carcinomas. *Science* 244, 217–221

81. Hermsen, M., Postma, C., Baak, J., Weiss, M., Rapallo, A., Sciutto, A., Roemen, G., Arends, J. W., Williams, R., Giaretti, W., de Goeij, A., and Meijer, G. (2002) Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology* 123, 1109–1119

82. Colorectal Cancer - Staging of Colorectal Cancer Colorectal Cancer - Staging of Colorectal Cancer. *hopkinscoloncancercenter.org*,

83. Fodde, R., Smits, R., and Clevers, H. (2001) APC, Signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer* 1, 55–67

84. Bos, J. L., Fearon, E. R., Hamilton, S. R., Vries, M. V. de, van Boom, J. H., van der Eb, A. J., and Vogelstein, B. (1987) Prevalence of ras gene mutations in human colorectal cancers. *Nature* 327, 293–297

85. Forrester, K., Almoguera, C., Han, K., Grizzle, W. E., and Perucho, M. (1987) Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature* 327, 298–303

86. Nagasaka, T. (2004) Colorectal Cancer With Mutation in BRAF, KRAS, and Wild-Type With Respect to Both Oncogenes Showing Different Patterns of DNA Methylation. *Journal of Clinical Oncology* 22, 4584–4594

87. Zha, S., Yegnasubramanian, V., Nelson, W. G., Isaacs, W. B., and De Marzo, A. M. (2004) Cyclooxygenases in cancer: progress and perspective. *Cancer Letters* 215, 1–20

88. CE, E., RJ, C., A, R., FM, G., S, F., and RN, D. (1994) Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas. *Gastroenterology* 107, 1183–1188

89. Maekawa, M., Sugano, K., Sano, H., Miyazaki, S., Ushiama, M., Fujita, S., Gotoda, T., Yokota, T., Ohkura, H., Kakizoe, T., and Sekiya, T. (1998) Increased Expression of Cyclooxygenase-2 to -1 in Human Colorectal Cancers and Adenomas, but not in Hyperplastic Polyps. *Japanese Journal of Clinical Oncology* 28, 421–426

90. Sabates-Bellver, J., Van der Flier, L. G., de Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M. A., Bujnicki, J. M., Menigatti, M., Luz, J., Ranalli, T. V., Gomes, V., Pastorelli, A., Faggiani, R., Anti, M., Jiricny, J., Clevers, H., and Marra, G. (2007) Transcriptome profile of human colorectal adenomas. *Mol. Cancer Res.* 5, 1263–1275

91. Maglietta, R., Liuzzi, V. C., Cattaneo, E., Laczko, E., Piepoli, A., Panza, A., Carella, M., Palumbo, O., Staiano, T., Buffoli, F., Andriulli, A., Marra, G., and

Ancona, N. (2012) Molecular pathways undergoing dramatic transcriptomic changes during tumor development in the human colon. *BMC Cancer* 12, 608

92. Vonlanthen, J., Okoniewski, M. J., Menigatti, M., Cattaneo, E., Pellegrini-Ochsner, D., Haider, R., Jiricny, J., Staiano, T., Buffoli, F., and Marra, G. (2014) A comprehensive look at transcription factor gene expression changes in colorectal adenomas. *BMC Cancer* 14, 46

93. Nambiar, P. R., Gupta, R. R., and Misra, V. (2010) An “Omics” based survey of human colon cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 693, 3–18

94. Goelz, S. E., Vogelstein, B., Hamilton, S. R., and Feinberg, A. P. (1985) Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* 228, 187–190

95. Feinberg, A. P., Gehrke, C. W., Kuo, K. C., and Ehrlich, M. (1988) Reduced Genomic 5-Methylcytosine Content in Human Colonic Neoplasia. *Cancer Res*,

96. Silverman, A. L., Park, J. G., Hamilton, S. R., Gazdar, A. F., Luk, G. D., and Baylin, S. B. (1989) Abnormal methylation of the calcitonin gene in human colonic neoplasms. *Cancer Res.* 49, 3468–3473

97. Suzuki, H., Watkins, D. N., Jair, K.-W., Schuebel, K. E., Markowitz, S. D., Dong Chen, W., Pretlow, T. P., Yang, B., Akiyama, Y., van Engeland, M., Toyota, M., Tokino, T., Hinoda, Y., Imai, K., Herman, J. G., and Baylin, S. B. (2004) Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat. Genet.* 36, 417–422

98. WARBURG, O. (1956) On the origin of cancer cells. *Science* 123, 309–314

99. Koppenol, W. H., Bounds, P. L., and Dang, C. V. (2011) Otto Warburg's contributions to current concepts of cancer metabolism.

100. Mullen, A. R., and Deberardinis, R. J. (2012) Genetically-defined metabolic reprogramming in cancer. *Trends Endocrinol. Metab*,

101. Ward, P. (2012) ScienceDirect.com - Cancer Cell - Metabolic Reprogramming: A Cancer Hallmark Even Warburg Did Not Anticipate. *Cancer Cell*,

102. Wang, H., Tso, V., Wong, C., Sadowski, D., and Fedorak, R. N. (2014) Development and validation of a highly sensitive urine-based test to identify patients with colonic adenomatous polyps. *Clin Transl Gastroenterol* 5, e54

103. Nugent, J. L., McCoy, A. N., Addamo, C. J., Jia, W., Sandler, R. S., and Keku, T. O. (2014) Altered Tissue Metabolites Correlate with Microbial Dysbiosis in Colorectal Adenomas. *J Proteome Res* 13, 1921–1929

104. Shrubsole, M. J., Cai, Q., Wen, W., Milne, G., Smalley, W. E., Chen, Z., Ness, R. M., and Zheng, W. (2012) Urinary Prostaglandin E2 Metabolite and Risk for Colorectal Adenoma. *Cancer Prev Res (Phila)* 5, 336–342

105. Patterson, S. D., and Aebersold, R. H. (2003) Proteomics: the first decade and beyond. *Nat. Genet.* 33, 311–323

106. Stulik, J., Hernychova, L., Porkertova, S., Knizek, J., Macela, A., Bures, J., Jandik, P., Langridge, J. I., and Jungblut, P. R. (2001) Proteome study of colorectal carcinogenesis. *Electrophoresis* 22, 3019–3025

107. Alfonso, P., Canamero, M., Fernandez-Carbonie, F., Nunez, A., and Casal, J. I. (2008) Proteome Analysis of Membrane Fractions in Colorectal Carcinomas by Using 2D-DIGE Saturation Labeling. *J Proteome Res* 7, 4247–4255

108. Mazzanti, R., Solazzo, M., Fantappie, O., Elfering, S., Pantaleo, P., Bechi, P., Cianchi, F., Ettl, A., and Giulivi, C. (2006) Differential expression proteomics of human colon cancer. *American Journal of Physiology-Gastrointestinal and Liver Physiology* 290, G1329–G1338

109. Polley, A. C. J., Mulholland, F., Pin, C., Williams, E. A., Bradburn, D. M., Mills, S. J., Mathers, J. C., and Johnson, I. T. (2006) Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia. *Cancer Res.* 66, 6553–6562

110. Patel, B. B., Li, X.-M., Dixon, M. P., Blagoi, E. L., Seeholzer, S. H., Chen, Y., Miller, C. G., He, Y. A., Tetruashvily, M., Chaudhry, A. H., Ke, E., Xie, J., Cooper, H., Bellacosa, A., Clapper, M. L., Boman, B. M., Zhang, T., Litwin, S., Ross, E. A., Conrad, P., Crowell, J. A., Kopelovich, L., Knudson, A., and Yeung, A. T. (2007) Searchable high-resolution 2D gel proteome of the human colon crypt. *J Proteome Res* 6, 2232–2238

111. Alfonso, P., Nunez, A., Madoz-Gurpide, J., Lombardia, L., Sánchez, L., and Casal, J. I. (2005) Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis. *Proteomics* 5, 2602–2611

112. Rho, J.-H., Qin, S., Wang, J. Y., and Roehrl, M. H. A. (2008) Proteomic Expression Analysis of Surgical Human Colorectal Cancer Tissues: Up-Regulation of PSB7, PRDX1, and SRP9 and Hypoxic Adaptation in Cancer. *J Proteome Res* 7, 2959–2972

113. F Lam, F., Jankova, L., Dent, O. F., Molloy, M. P., Kwun, S. Y., Clarke, C., Chapuis, P., Robertson, G., Beale, P., Clarke, S., Bokey, E. L., and Chan, C. (2010) Identification of distinctive protein expression patterns in colorectal adenoma. *Prot. Clin. Appl.* 4, 60–70

114. Domon, B., and Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* 28, 710–721

115. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* 9, 555–566

116. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*,

117. Domon, B., and Aebersold, R. (2006) Mass Spectrometry and Protein Analysis. *Science* 312, 212–217

118. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics. *Chemical reviews*,

119. Wu, W. W., Wang, G., Baek, S. J., and Shen, R.-F. (2006) Comparative Study of Three Proteomic Quantitative Methods, DIGE, cICAT, and iTRAQ, Using 2D Gel- or LC−MALDI TOF/TOF. *J Proteome Res* 5, 651–658

120. Bensimon, A., Heck, A. J. R., and Aebersold, R. (2012) Mass Spectrometry-Based Proteomics and Network Biology. *Annu. Rev. Biochem.* 81, 379–405

121. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 389, 1017–1031

122. Jimenez, C. R., Knol, J. C., Meijer, G. A., and Fijneman, R. J. A. (2010) Proteomics of colorectal cancer: Overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *Journal of Proteomics* 73, 1873–1895

123. de Wit, M., Fijneman, R. J. A., Verheul, H. M. W., Meijer, G. A., and Jimenez, C. R. (2012) Proteomics in colorectal cancer translational research: Biomarker discovery for clinical applications. *Clin. Biochem*,

124. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 4, 222

125. Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L., and Park, B. K. (2009) Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics☆. *Journal of Chromatography B* 877, 1229–1239

126. Hüttenhain, R., Malmström, J., Picotti, P., and Aebersold, R. (2009) Perspectives of targeted mass spectrometry for protein biomarker verification. *Current Opinion in Chemical Biology* 13, 518–525

127. Hörth, P., Miller, C. A., Preckel, T., and Wenz, C. (2006) Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol. Cell Proteomics* 5, 1968–1974

128. Grossmann, J., Roschitzki, B., Panse, C., Fortes, C., Barkow-Oesterreicher, S., Rutishauser, D., and Schlapbach, R. (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *Journal of Proteomics* 73, 1740–1746

129. Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C., and Scrivens, J. H. (2009) A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* 8, 3752–3759

130. Mueller, L. N., Rinner, O., Schmidt, A., Letarte, S., Bodenmiller, B., Brusniak, M.-Y., Vitek, O., Aebersold, R., and Mueller, M. (2007) SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 7, 3470–3480

131. Braisted, J. C., Kuntumalla, S., Vogel, C., Marcotte, E. M., Rodrigues, A. R., Wang, R., Huang, S.-T., Ferlanti, E. S., Saeed, A. I., Fleischmann, R. D., Peterson, S. N., and Pieper, R. (2008) The APEX Quantitative Proteomics Tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* 9,

132. Ong, S.-E., and Mann, M. (2007) in *Quantitative Proteomics by Mass Spectrometry*, Methods in Molecular Biology. (Humana Press, Totowa, NJ), pp 37–52.

133. Wiese, S., Reidegeld, K. A., Meyer, H. E., and Warscheid, B. (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7, 340–350

134. Han, D. K., Eng, J., Zhou, H. L., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* 19, 946–951

135. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., and Coon, J. J. (2012) Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell Proteomics* 11, 1475–1488

136. Picotti, P., Rinner, O., Stallmach, R., and Dautel, F. (2009) High-throughput generation of selected reaction-monitoring assays for proteins and proteomes : Abstract : Nature Methods. *Nat. Methods*,

137. Collins, B. C., Gillet, L. C., Rosenberger, G., Roest, H. L., Vichalkovski, A., Gstaiger, M., and Aebersold, R. (2013) Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* 10, 1246–+

138. Liu, Y., Hüttenhain, R., Surinova, S., Gillet, L. C. J., Mouritsen, J., Brunner, R., Navarro, P., and Aebersold, R. (2013) Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *Proteomics* 13, 1247–1256

139. Cattaneo, E., Laczko, E., Buffoli, F., Zorzi, F., Bianco, M. A., Menigatti, M., Bartosova, Z., Haider, R., Helmchen, B., Sabates-Bellver, J., Tiwari, A., Jiricny, J., and Marra, G. (2011) Preinvasive colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid vs. nonpolypoid). *EMBO Mol Med* 3, 334–347

4 Aims of Study

Phenotypic variations in the mucosa of the large bowel, as displayed in endoscopy, define the different stages associated with colorectal cancer. Colorectal cancer is prevalent in advance-aged individuals and is one of the major causes of cancer-related deaths worldwide. The early identification and resection of precancerous lesions detected during endoscopic screening remains the most effective way to reduce the incidence of the disease. Molecular research on colorectal tumors has contributed immensely to present day knowledge on the development of various cancers. More recently, the application of new techniques in the field of proteomics, transcriptomics, and metabolomics, has helped unravel novel gene interactions and pathways involved in colorectal tumorigenesis. Limited information exists on global proteome alterations pertinent to precancerous lesions of the large bowel. Research on colorectal precancerous lesions is important to shed more light on the biology of these benign tumors and the molecular mechanisms that control their transformation to malignant lesions. Furthermore, protein markers for these lesions could be putative biomarkers for the early detection of colorectal carcinogenesis.

The first aim of my study was to optimize a quantitative shotgun proteomics workflow to identify and relatively quantify protein expression in a large series of human colorectal adenomas and matched normal mucosa samples (discovery phase). And in particular, to apply this quantitative approach to identify proteins and biological processes altered in adenomas compared to the normal mucosa. The proteome of 6 human colon epithelial cell lines (five cancerous and one non-cancerous cell line) was also investigated by quantitative proteomics. Since colorectal tumorigenesis originates on the epithelial mucosa, and human biopsies contain epithelial and stromal cells, proteome similarity between cell lines and tissues was hypothesized to be indicative of protein alterations specific to the colon epithelium. An iTRAQ-8plex shotgun workflow was optimized to achieve this. Additionally, we investigated the correlation of proteome and gene expression data from studies on human colorectal precancerous lesions using statistical tools. For this, proteomic data from my shotgun study was compared with transcriptomic data from studies (139) earlier reported on a different series of adenoma and adjacent normal mucosa tissues.

The final goal of my study was to apply a targeted proteomic approach to validate putative tumor markers for colorectal adenomas in a large number of adenoma and

adenocarcinoma tissues (with normal mucosa pairs). High-throughput SRM assays were developed and optimized to detect and reproducibly quantify selected proteins candidates. These protein candidates selected from my shotgun study and from the transcriptomic study of Cattaneo et al., (139) were potential tumor biomarkers for the early stages of colorectal tumorigenesis.

5 Results

5.1 Quantitative shotgun proteomics reveals early markers of colorectal carcinogenesis

The first aim of my study was to optimize a quantitative shotgun proteomics workflow to comprehensively investigate protein expression changes that arise during the formation of colorectal adenomas and to identify putative tumor markers for these lesions.

5.1.1 INTRODUCTION

Colorectal cancer ranks third among the world's high-incidence cancers and is a leading cause of cancer-related death among older adults (1, 2). In the United States alone, projections for 2013 include 102,480 new cases and 50,830 deaths (2). Cancerogenesis in the large bowel begins with the transformation of the epithelial cell lining of the gut. Molecular alterations, mainly involving the WNT signaling pathway, render these cells hyperproliferative, and they form benign adenomatous tumors. The neoplasms are initially noninvasive (3, 4), and the vast majority remain that way. But as genetic and epigenetic anomalies continue to accumulate, the tumor cells' capacity for invasion and destruction of surrounding tissues increases. At some point, this process drives certain adenomas into the realm of frank malignancy, transforming them into adenocarcinomas.

Early diagnosis of colorectal tumors has been greatly facilitated by screening methods based on fecal analysis or colonoscopy, but both approaches have limitations (5-9). Better understanding of the molecular mechanisms underlying large bowel tumorigenesis could improve our chances of detecting these lesions in the adenomatous or localized adenocarcinomatous stage, when the chances of successful treatment are greater. Promising results for the detection and validation of potential cancer biomarkers are emerging from proteomic studies of cancer development (10). Compared with older gel electrophoresis-based approaches, shotgun proteomic methods, particularly those that include pre-MS OFFGEL electrophoretic peptide fractionation (11), enhance the sensitivity, robustness, and reproducibility of these studies (12) and expand the proteome coverage to include proteins that are less abundantly expressed (13-16). Furthermore, with the aid of isobaric-tag peptide labeling strategies, MS can also be used for relative quantification of protein expression levels within a series of multiple human tissue samples (12, 17-19).

Thus far, only a few MS-based proteomic studies have examined human colorectal adenomas (reviewed in refs. (9, 20)). We therefore decided to explore the proteome of a relatively large series of these precancerous lesions (each with a paired sample of normal colon mucosa) using quantitative shotgun MS with the widely used iTRAQ (isobaric tags for relative and absolute quantitation) peptide labeling technique (21, 22) and OFFGEL fractionation. Adenoma-related protein expression variations specific to the epithelial compartment of these lesions were identified with a novel approach, which involved comparing the human tissue proteome with that of colon epithelial cell lines. The results of these studies revealed several protein expression

changes previously documented only in advanced colorectal cancers. They also disclosed several novel changes with potentially important roles in early-stage large bowel tumorigenesis, including the marked upregulation of a key enzyme in the polyol pathway.

5.1.2 EXPERIMENTAL PROCEDURES

5.1.2.1 Human tissue samples and cell lines

Human colorectal tissues were prospectively collected from patients undergoing colonoscopy in the Istituti Ospitalieri of Cremona, Italy. Local ethics committee approval was obtained, and tissues were used in accordance with the Declaration of Helsinki. Each donor provided written informed consent to sample collection, data analysis, and publication of the findings. Progressive numbers were assigned to each patient to protect human confidentiality. The series comprised 30 colorectal adenomas, each with a paired sample of normal mucosa from the same colon segment, >2 cm from the lesion. Tissues were collected endoscopically, promptly frozen in liquid nitrogen, and stored at -80 ^{o}C.

Five colorectal cancer cell lines (HT29, Caco2, CX1, SW480, SW620) were obtained from the Zurich Cancer Network's Cell Line Repository. All had been recently purchased from the American Tissue Culture Collection (Teddington, UK) and were certified to be mycoplasma infection-free. Caco2 and CX1 cells were cultured in Dulbecco's Modified Eagle Medium, HT29 cells in McCoy's medium, and SW480 and SW620 cells in RPMI 1640 medium supplemented with 10% fetal bovine serum, L-glutamine, and 1% penicillin-streptomycin (Sigma, St Louis, Missouri, USA). The recently established line of immortalized human colon epithelial cells (HCEC) was obtained from J. W. Shay and grown as described elsewhere (23).

5.1.2.2 Protein extraction from tissues and cell lines

For MS studies, frozen tissue samples were quickly weighed and homogenized on ice (1 min grinding, 1 min on ice, 1 min grinding) in a Wheaton glass borosilicate grinder containing a solution of 100 mM triethylammonium bicarbonate (Sigma, St Louis, MO, USA), 1X Complete EDTA-free Protease Inhibitor Cocktail (Roche, Mannheim, Germany), 1 M urea, 5 mM β-glycerophosphate disodium salt hydrate, 1 mM sodium orthovanadate, and 5 mM sodium fluoride (Sigma). The efficiency of cell lysis was microscopically confirmed. The homogenates were then sonicated with a Bioruptor (Diagenode, Denville, NJ, USA) (high power, five 10"/10" ON/OFF cycles) and centrifuged (16,000g for 5 min at 4 ^{o}C). The supernatant containing the proteins was collected and stored at -80 ^{o}C.

Cells (grown to >80% confluence in 15 cm^2 dishes) were washed in PBS, covered with 250 µL of the buffer used for tissue sample homogenization (see above), detached

from the dish with a cell scraper, and homogenized (25 passages through a 25G needle). The efficiency of cell lysis was microscopically confirmed. Sonication and centrifugation were repeated, as described above, and the protein concentration was determined by Bradford assay. Prior to MS analysis, a 5-µg sample of each protein extract was subjected to 1-dimensional gel electrophoresis on a 12% bisacrylamide gel to assess protein integrity and extraction protocol reproducibility. The entire proteomic workflow, from tissue/cell processing to statistical analysis, is summarized in Figure 1 and described in detail in the next five sections of Experimental procedures.

For sorbitol dehydrogenase (SORD) assays (see below), >80% confluent cells were washed in PBS and covered with a solution consisting of 100 mM triethanolamine (Sigma) and 1X Complete EDTA-free Protease Inhibitor Cocktail (Roche). (A simple buffer was used to reduce the risk of introducing anti-enzymatic substances into our extract.) Cells were then scraped and homogenized with 25 passages through a 25G needle. Tissue samples were weighed and homogenized in a Wheaton glass borosilicate grinder containing the buffer described above. After centrifugation (16,000 *g*, 4 ^{o}C, 5 min), the supernatant was aliquoted and stored at -80 ^{o}C. Protein concentration was measured by Bradford assay.

5.1.2.3 Protein digestion and iTRAQ 8-plex labeling

iTRAQ 8-plex experiments were performed to analyze tissue extracts (10 experiments) and cell-line extracts (1 experiment) (Figure 1). Labeling efficiency and relative quantitation accuracy were assessed with the aid of two reference protein extract mixtures: one for tissue samples (pooled extracts from 3 normal tissues and 3 adenomas) and one for cell lines (pooled aliquots of each of the six cell line extracts). Fifty micrograms of protein per sample were used for each iTRAQ channel. Tryptic digestion (10% w/w, sequencing-grade modified trypsin; Promega, Madison, WI, USA) and iTRAQ 8-plex labeling (AB Sciex, Framingham, MA, USA) were performed according to manufacturers' instructions (2.5-hour incubation of samples with iTRAQ labels). For tissue experiments, two iTRAQ labels, 113 and 114, were chosen for the reference mixture, while labels 115/116, 117/118, and 119/121 were used for the 3 pairs of normal/adenomatous tissues included in each experiment. For the cell line experiment, labels 113 and114 were used for the reference mixture, and labels 115-121 represented HCEC, HT29, Caco2, CX1, SW480, and SW620 cells, respectively (Figure 1). After iTRAQ labeling, the samples (for each experiment) were combined, desalted on 500 mg SepPak C18 columns (Millipore, Billerica, MA, USA), dried in a

SpeedVac concentrator (Thermo Scientific, NC, USA), and subjected to peptide fractionation.

5.1.2.4 OFFGEL electrophoresis

Peptide fractionation was performed according to the manufacturer's protocols with an Agilent 3100 OFFGEL fractionator and 12-well OFFGEL kit (both from Agilent Technologies, Santa Clara, CA, USA). Briefly, samples were resolubilized in 1.8 mL of 1x OFFGEL peptide stock solution containing carrier ampholytes (pH range 3–10), loaded into the wells (150 µL per well), and focused until 20 kV/h was reached with a maximum current of 50 µA. For each experiment, 12 fractions were collected. A 15-µL aliquot of each fraction was acidified with 1.5 µL of a 50% acetonitrile / 1% trifluoroacetic acid solution, desalted using ZipTip C18 (Millipore, Billerica, MA, USA), dried, resolubilized in 15 µL of a 0.1% formic acid / 3% acetonitrile solution, and analyzed with MS.

5.1.2.5 Liquid Chromatography and Mass Spectrometry

Peptide samples (4 µL) were analyzed on an LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to a nano-HPLC system (Eksigent Technologies, Dublin, CA, USA). The solvent compositions were 0.2% formic acid and 1% acetonitrile for channel A and 0.2% formic acid and 80% acetonitrile for channel B. Peptides were loaded onto an in-house-made tip column (75 µm × 80 mm) packed with reverse-phase C18 material (AQ, 3 µm 200 Å, Bischoff GmbH, Leonberg, Germany) and eluted (flow rate, 250 nL/min; solvent B gradient: from 3% to 30% in 62 min, from 30% to 45% in 70 min, and 45% to 97% in 75 min).

Full-scan MS spectra (300−1700 m/z) were acquired at a resolution setting of 30 000 at 400 m/z after accumulation to a target value of 1×10^6. For the eight most intense signals per cycle above a threshold of 1000, both CID (collision-induced dissociation) and HCD (higher-energy collisional dissociation) spectra were acquired in a data-dependent manner (Figure 1). CID scans were recorded in the ion-trap (settings: normalized collision energy, 35; maximum injection time, 50 ms; automatic gain control [AGC], 1×10^4 ions). For the HCD scans, spectra were recorded at a resolution setting of 7500 at 400m/z (normalized collision energy, 52; maximum injection time, 125 ms; AGC, 5×10^4 ions). Charge state screening was enabled, and singly charged states were rejected. Precursor masses previously selected for MS/MS were excluded from further selection for 60 s, and the exclusion window was set at 10 ppm. The maximum number of entries in the exclusion list was set at 500. All samples

were analyzed in duplicate, and precursors selected in the first run were excluded from fragmentation in the second run. The exclusion list was set on a time window of 4 minutes and mass width of 10 ppm. Spectra were acquired using internal lock mass calibration on m/z 429.088735 and 445.120025.

5.1.2.6 Peak list generation and database search

As depicted in Figure 1, Mascot Distiller 2.4.3.3 (Matrix Science, Boston, MA, USA) was used to generate Mascot generic format (MGF) peak lists. De-isotoping and peak picking were not performed between 112.5 and 121.5 m/z (the range containing iTRAQ reporter ions), and the HCD and CID spectra were merged by summing. For each of the 11 experiments, the corresponding 24 MGF peak lists were concatenated and searched, with the aid of the Mascot Server 2.3.02 (Matrix Science), against a forward UniProtKB/Swiss-Prot database for human proteins concatenated to a reversed decoyed FASTA database. The concatenated database contained a total of 147,438 proteins with accessions in Gene Ontology-compatible format and 260 common MS contaminants (NCBI taxonomy ID 9606, release date 2011-12-13).

Methylthio (C), iTRAQ 8-plex labeling at N-terminal and lysine were set as fixed modifications, while variable modifications consisted of methionine oxidation and iTRAQ 8-plex labeling of tyrosine. We used the iTRAQ 8-plex-vs114 (Applied Biosystems) quantitation method. The isotope and impurity correction factors used for each iTRAQ label were those provided by the manufacturer. Precursor and fragment tolerances were set at 10 ppm and 0.8 Da, respectively. Enzyme specificity was set to trypsin with an allowance of up to 1 missed cleavage. Using Mascot internal export scripts, we transformed Mascot DAT files into XML files and parsed them with in-house scripts so that peptide sequences, scores, and intensities of the individual reporter ion channels were reported. Confidently identified and quantified peptides were selected with the following filters: rank 1 (best spectra assignment); ion score: > 15; and presence of iTRAQ intensity values for reporter channels 113 and 114.

5.1.2.7 Quantification of relative protein abundance

(these steps are described in the boxes of the lower half of Figure 1)

Peptide reporter channel intensities were summed for each protein individually using R-scripts. Ratios were built from summed channels (113/114 to 121/114) for all proteins identified in each iTRAQ experiment. False discovery rates (FDRs) (24) were determined at the spectrum, peptide, and protein levels. The results of individual experiments were then merged into one matrix, which was used for statistical analysis

in R and Perseus (Version 1.2.7.4). All proteins identified with the same peptide(s) were grouped into families, each of which was identified by a unique protein family number. Ratios of the intensity of each ion channel to that of 114 were converted to base 2 logarithmic values and normalized respectively on the median (which was set at 0), resulting in ratios that followed a Gaussian distribution. Proteins identified on the basis of the same peptide(s) shared the same family number and were represented once in our statistical analysis. The paired t-test was used to compare the expression of a given protein in each adenoma and that found in the corresponding sample of normal mucosa. To correct for multiple comparisons, the FDR was controlled with the Benjamini-Hochberg procedure. The average protein-expression fold change in adenomas, compared with the normal mucosa, was then calculated. For this, median normalized ratios for all proteins in each paired adenoma-normal mucosa sample were deconvoluted of the reference standard effects (114) to compute the adenoma vs. normal mucosa ratio per protein (deconvoluted fold change, i.e.[116/114] / [115/114] = [116/115]) and the mean fold change per protein in all tissue pairs. The Mascot emPAI value for all proteins were included in XML exports for each experiment. Thereafter, the mean Mascot emPAI value was calculated for all proteins.

5.1.2.8 Functional annotation of proteins

Gene ontology (GO) annotations and GO terms for proteins in the UniProt/SwissProt database were sourced from http://www.ebi.ac.uk/. The Scaffold program (Version 3) was used to identify the cellular localizations and biological processes most represented in lists of proteins quantified in tissues and cell lines. The topGO Bioconductor software package in R was used to identify and screen for GO biological process categories displaying enrichment for proteins that were differentially regulated in adenomas (vs. normal mucosa) (25). First, we prepared a "Universe" comprising all the proteins quantified in our study, each matched to GO terms and annotations. This served as the "Background." The "Foreground" consisted of the list of significantly dysregulated proteins. The most significant GO terms were scored with the Eliminating Genes (elim) method (25).

5.1.2.9 Western Blotting

Proteins were separated on a 10% SDS-polyacrylamide gel and transferred to a hydrophobic polyvinylidene difluoride membrane (GE Healthcare, Amersham Hybond-P PVDF membrane, Pittsburgh, PA, USA) according to standard protocols (26). After 1 h of blocking with 5% non-fat dry milk in TBS with 1% Tween 20 (milk-TBST),

membranes were incubated overnight with the primary antibody (anti-SORD [HPA040260 Sigma]; anti-aldose reductase, AKR1B1 [GTX113381 GeneTex]; anti-ketohexokinase, KHK [GTX109591 GeneTex]) diluted 1:1000 in milk-TBST, washed once with milk-TBST (20 min) and twice with TBST (20 min). After 1 h of incubation in horseradish peroxidase-conjugated secondary antibody (anti-rabbit IgG, GE Healthcare) diluted 1:5000 in milk-TBST, membranes were washed once with milk (20 min) and twice with TBST (20 min). Enhanced chemiluminescence (ECL, Amersham Pharmacia Biotech, Cat. No. RPN2106) was used to detect immunoreactive proteins.

5.1.2.10 Immunostaining of cells and tissues

HT29 and HCEC cells were seeded (3 x 10^5 per well) on 22 x 22 mm cover slips in 6-well plates and grown under standard conditions until cells reached 70-80% confluence. Cells were then washed twice with PBS and fixed in ethanol:methanol solution (50:50) for 10 min at room temperature. Fixed cells were permeabilized (10 min with 0.25% Triton X-100), blocked (30 min in 10% goat serum [X0907, Dako, Glostrup, Denmark]), and incubated with primary antibody (rabbit polyclonal anti-SORD, HPA040260, Sigma, 1:100) for 18 h at 4 ^{o}C. After three washes with PBS, the cells were incubated for 1 h with secondary antibody conjugated to polymer-HRP anti-rabbit (Dako, EnVision+ System- HRP; Cat. No. K4010). They were then washed three times in PBS and incubated for 15 min in the substrate-chromogen, 3,3-diaminobenzidine tetrahydrochloride (DAB, Dako, EnVision+ System- HRP; Cat. No. K4010). Cells were washed quickly with PBS and mounted on slides (EUKITT, O. Kindler, GmbH, Freiburg, Germany) for light microscopy (Leica Microsystems GmbH, Wetzlar, Germany). Images were examined and recorded with Leica Application Suite (V3.3.0) software.

Sorbitol dehydrogenase immunohistochemistry was performed as previously described (27). Tissue sections (normal colon and ileum, colorectal adenomas and adenocarcinomas) were incubated for 24 h at 4 ^{o}C with primary antibody (rabbit polyclonal anti-SORD, HPA040261, Sigma) at a 1:100 dilution.

5.1.2.11 Measurement of SORD activity

Total protein was extracted from cell lines and tissues as described above. SORD catalyzes the reversible conversion of D-sorbitol to D-fructose, with β-NADH as a cofactor:

D-fructose + β-NADH <----------> D-sorbitol + β-NAD

SORD

SORD activity was quantified via continuous spectrophotometric rate measurement of the β-NAD formation rate (temperature 25 ^{o}C, pH 7.6, A340nm, light path of 1 cm) in a Cary 50 Scan UV-visible spectrophotometer using the Cary Kinetics Application (both from Varian Inc., Palo Alto, CA, USA) (28). The final reagent concentrations in a 1-mL cuvette were as follows: 78.33 mM triethonolamine, 183 mM D-fructose, 0.21 mM β-NADH, 0.033% (w/v) BSA. The absorbance reading was recorded when the enzyme was added. One unit of enzyme activity was defined as the amount of enzyme required per minute to convert 1.0 µmole of D-fructose to D-sorbitol at pH 7.6 at 25 ^{o}C. A mixture of reagents plus recombinant SORD was used as the positive control; negative controls consisted of the same reagent mixture with no recombinant SORD, with recombinant SORD but no D-fructose, or with recombinant SORD but no β-NADH.

5.1.2.12 Extraction and quantification of intracellular metabolites by targeted gas chromatography-coupled MS

Frozen tissue (50 to 100 mg) was homogenized in 250 µl ice-cold 80% methanol using a glass borosilicate grinder from Wheaton. The homogenate was microscopically examined to ensure that it was cell-free and then transferred to Eppendorf vials and left on ice for 15 min to ensure efficient protein precipitation. After centrifugation (15,000g for 3 min at 0 ^{o}C), the supernatant was snap-frozen and stored at -80 ^{o}C, while the protein content of the pellet was determined by the Bradford method.

For gas chromatography-coupled MS (GC-TOF-MS) analyses, 10 µL of supernatant was transferred to a 1.5-mL Eppendorf tube, and internal standards ($^{13}C^{1}$-sorbitol, $^{13}C^{1}$-fructose, and $^{13}C^{1}$-glucose—1.2 pmoles of each) were added. The samples were then dried overnight in a vacuum centrifuge (Concentrator 5301, Eppendorf AG, Germany). Methoxyamine hydrochloride and N-methyl-N-(trimethylsilyl) trifluoroecetamide were used as derivatization reagents (29).

The derivatized metabolites and internal standards were subjected to GC-TOF-MS (GC 7890A, Agilent Technologies, Santa Clara, CA, USA; GCT Premier Micromass, Waters, Manchester, UK) with a Rxi-5Sil MS Integra-Guard column (length: 30 meters, internal diameter: 0.25 mm) and a film thickness of 0.25 µm (Restek, USA). One microliter of each derivatized sample was injected in splitless mode on a baffled glass liner and transferred to the capillary column by rapid heating of the liner from 50 ^{o}C to 250 ^{o}C at a rate of 12 ^{o}C/sec. For the separation of the metabolites, helium was used at a flow rate of 1 ml/min and, after an initial hold time of 2 min, a temperature gradient from 80 ^{o}C to 320 ^{o}C (rate 8 ^{o}C/min) was applied. The TOF-MS was set to acquire

centroided standard electron ionization mass spectra over a range of 50 to 600 m/z at a rate of 3 spectra/sec. The GC-MS transfer line was heated to 280 ^{o}C. Dynamic range enhancement was activated. C_6ClF_5 was used as lock mass compound.

The MassLynx and QuanLynx programs (Waters, UK) were used to review and analyze the acquired data. The absolute concentrations of D-sorbitol, α and β D-fructose, and α and β D-glucose were calculated on the basis of the ratio of the intensity of specific fragments originating from the unlabeled compound to that of the added labeled analogue (internal standard). These concentrations were used to estimate intracellular levels per milligram of tissue (adenoma vs. normal mucosa). The relative concentration of lactate was estimated from the ratio of the intensity of specific fragments originating from the unlabelled compound and that of the added $^{13}C_1$-sorbitol (internal standard).

5.1.3 RESULTS

5.1.3.1 Proteomic analysis of human colorectal tissues and colon cell lines.

We used a quantitative-MS-based discovery strategy to explore the proteome of human colorectal tissues and colon cell lines (normal and neoplastic). The characteristics of the precancerous colorectal lesions are listed in Table 1. Protein extracts from these tumors and their paired samples of normal mucosa (total, 60 samples) were analyzed using iTRAQ LC-MS/MS and the workflow described in Figure 1.

Table 1. Characteristics of the precancerous lesions included in the study

Patient number	Age	Sex	Colon segment involved	Maximum lesion diameter (mm)	Paris classification #	Pit pattern classification °	Microscopic appearance	Highest degree of dysplasia in the lesion ▪	No. of lesions at study colonoscopy ∞	No. of previously excised lesions ‡
1	77	M	S	25	IIa+IIc	IIIs - IIIL	TA	LGD	1	0
2	73	F	A	25	IIa+IIc	IIIs-IIIL	TA	LGD	1	2
3	59	M	T	30	IIa+IIc	IIIs + IIIL	TA	LGD	1	0
4	73	F	R	50	Is	IV	VA	LGD	1	0
5	74	M	R	40	Is	IV	VA	HGD	2	1
6	77	M	C	25	IIa	IIIL	VA	LGD	1	0
7	80	M	A	40	IIa	IIIL	TVA	LGD	1	1
8	82	M	A	15	IIa	IIIL	VA	LGD	2	0
9	73	F	S	20	Ip	IV	TVA	LGD	1	0
10	70	F	C	25	IIa	IIIL	TVA	LGD	2	0
11	63	M	A	45	Is	IIIL-IV	TVA	LGD	0	0
12	68	M	A	30	IIa+Is	IIIL - IV	TVA	HGD	0	0
13	60	M	D	30	Is	IV - Vi	TVA	HGD	0	0
14	55	M	C	25	IIa+Is	IIIL-IV	SSA	LGD	0	0
15	70	M	A	15	Is	IV	TVA	LGD	7	0
16	85	F	S	25	Is+IIa	IV	TA	LGD	1	1
17	66	M	A	30	IIa	IIIL	TA	HGD	2	0
18	72	M	A	30	Is	IV	TVA	HGD	2	0
19	71	M	S	30	IIa	IIIL	TVA	LGD	2	0
20	59	M	R	60	Is	IV-Vi	TVA	HGD	1	0
21	78	M	A	50	Is	IV - Vi	TA	LGD	1	0
22	75	M	R	25	Is	IV - Vn	TVA	HGD	6	0
23	73	F	D	25	Is	IV	TA	LGD	1	0
24	69	F	R	90	Is+IIa	IV	TVA	LGD	1	0
25	75	M	T	18	IIa	IIIL	TA	LGD	1	0
26	61	M	A	40	Is+IIa	IV	TVA	LGD	20	0
27	76	M	S	30	Is	IV - Vi	TA	HGD	1	0
28	78	F	R	60	IIa+Is	IV	TVA	LGD	1	1
29	89	M	R	30	Is	IV	TA	LGD	3	0
30	75	M	A	50	Is	IV-Vn	TVA	HGD/cancer	7	0

Abbreviations: M, male; F, female; C, cecum; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum; TA, tubular adenoma; TVA, tubulovillous adenoma; VA, villous adenoma; SSA, sessile serrated adenoma; LGD, low-grade dysplasia; HGD, high-grade dysplasia.

Macroscopic appearance of neoplastic lesions was classified according to Paris Endoscopic Classification. The Paris Endoscopic Classification of Superficial Neoplastic Lesions. Gastrointest Endosc 2003;58(suppl.):S3-S27

° Morphological analysis of colon crypt patterns according to the Kudo classification. Kudo S, Rubio CA, Teixeira CR, et al. Pit pattern in colorectal neoplasia:endoscopic magnifying view. Endoscopy 2001;33;367-7

▪ Low-grade versus high-grade dysplasia as defined by the WHO classification of tumors of the digestive system, editorial and consensus conference in Lyon, France, November 6-9, 1999.IARC

∞ This number includes the lesion included in our proteomic study

The inclusion of two reference sample mixes allowed us to control for technical variability across the ten experiments on tissue samples since the reference sample was analyzed twice in each experiment. OFFGEL electrophoresis was used to obtain

highly reproducible, pI-based, in-solution separation of pooled iTRAQ-labeled peptides. Furthermore, for relative quantification of proteins using iTRAQ reporter ions, we adopted a stringent FDR for protein spectra matches (PSMs), and high-confidence peptides for protein quantification were selected only if the reporter ions (113 and 114) were quantified in the reference sample mix (iTRAQ reporter channels 113 and 114).

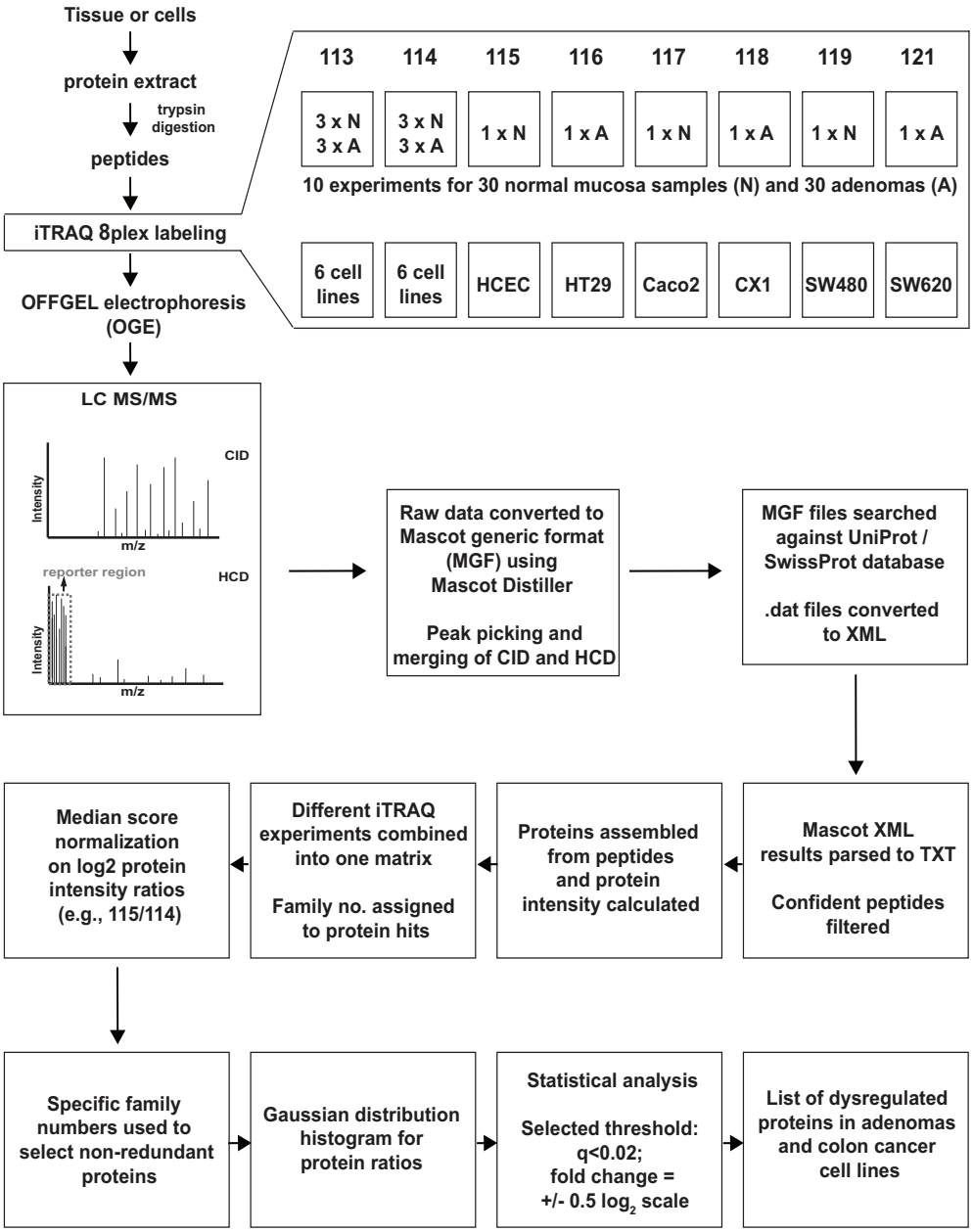


Figure 1. Project design and iTRAQ 8plex-labeling scheme

Project design and iTRAQ 8plex-labeling scheme. Sample preparation for shotgun MS/MS and important steps in the analysis of proteomic data for the detection of dysregulated proteins in adenomas and colon cancer cell lines. For each experiment on tissue samples, iTRAQ tags were assigned to a duplicate reference (two identical pools of normal and adenoma samples: 113 and 114, respectively), normal

tissues (115, 117, 119) and corresponding adenomas (116, 118, 121). The same pattern was repeated in all 10 experiments. In cell line experiments, two identical pools, each comprising all six cell lines, were used as reference (113 and 114), and each of the remaining six tags was used to label a single cell line. The data analysis flow chart depicted in this figure is described in Experimental procedures

The data set generated with this approach was large and complex, but we developed a simplified analytical method that allowed us to work with and merge the large data files generated after MS/MS (Figure 1). High-resolution MS/MS spectra acquired on the LTQ-Orbitrap Velos spectrometer after duplicate analysis of OFFGEL tissue sample fractions produced a total of 240 raw files (10 experiments, 120 fractions, 2 replicates).

Table 2. Summary of proteomics data

	Tissues (n = 60)		Cell lines (n = 6)	
	Total	FDR (%)	Total	FDR (%)
Peptide spectra matches	285,929	0.2	27,922	0.4
Peptides	37,184	0.9	11,266	0.5
Proteins *	10,452	1.5	5,056	1.1
Proteins #	4,325	-	2,017	-
Proteins ^	1,072	-	1,957	-

* Total number of proteins quantified in the 10 tissue experiments and the single experiment with cell lines.

Non-redundant protein families quantified in our dataset.

^ Non-redundant protein families quantified in all 60 tissues or in all 6 cell lines.

A total of 37,184 (FDR = 0.9%) unique tryptic peptides were confidently identified and quantified from 285,929 unique PSMs (FDR = 0.2%) (Table 2, Supplementary Table 1). Ten thousand four hundred fifty-two proteins (FDR = 1.5%) were assembled from the quantified peptides. Proteins that were indistinguishable by MS/MS (i.e., two or more proteins identified on the basis of the same peptide sequence; see Experimental procedures for details) were represented as a single family. The result was a total of 4325 non-redundant protein families, two-thirds (2865, 66%) of which were relatively quantified in at least 9 normal mucosa/adenoma pairs and 1072 (25%) in all 30 pairs (Table 2, Supplementary Table 1).

To verify the efficiency of iTRAQ protein labeling, we repeated the database search with Methylthio (C) set as a fixed modification and iTRAQ8plex (N-term), iTRAQ8plex (K), iTRAQ8plex (Y), and Oxidation (M) set as variable modifications. (All other search parameters were unchanged.) The assigned PSMs were filtered, as

described in Experimental procedures, and the average iTRAQ labeling efficiency achieved in each of the 10 tissue experiments was 96% (Supplementary Table 2).

To ascertain the efficacy of including a standard sample mix as a reference for normalization, we compared combined Gaussian plots of log_2-protein ratios of normal mucosa or adenoma samples with the respective reference channel per experiment (e.g., 115/114 vs. 113/114 for normal tissues, 116/114 vs. 113/114 for adenomas, see Figure 1). The ratios displayed normal distributions in all channels. For the reference channel (113/114), log_2-ratios were largely centered around 0, whereas the distribution of adenoma and normal channel log_2-ratios was broader and not always centered at 0 (data not shown).

Sample complexity is a common problem in the analysis of proteomic data from human colorectal tissues. It stems in part from contamination of the epithelial cell proteome by proteins from stromal cells (which were inevitably present in our specimens even though the endoscopic tissue sampling procedure we used yielded superficial specimens with consistently high epithelial contents). Microdissection can be utilized to isolate subpopulations of cells, but it can diminish the quantity and quality of the proteins, rendering them suboptimal for some types of proteomic analysis. To avoid this problem, we adopted a novel strategy for preliminary identification of the proteomic alterations that were most likely to involve the epithelial-cell component of the adenomas. The proteomic profiles of the colon tissues were compared with those of six colon epithelial cell lines (five colon cancer cell lines plus HCEC cells, to our knowledge, the only well-characterized line established from normal colorectal epithelium (23)). Changes in expression levels observed in adenomas (i.e., upregulation or downregulation with respect to normal mucosal levels) were presumed to be epithelial-cell-specific if similar changes were found in the colon cancer cell lines (relative to HCEC cells). After OFFGEL fractionation, duplicate MS analysis of iTRAQ-labeled peptides (24 fractions) from the six cell lines was performed in an LTQ-Orbitrap Velos mass spectrometer, and 11,266 peptides (FDR = 0.5%) were confidently identified and quantified from 27,922 unique PSMs (FDR = 0.4%) (Table 2, Supplementary Table 3). A total of 2017 non-redundant protein families (FDR = 1.1%) were identified and relatively quantified in cell lines; 1957 (97%) were present in all six cell lines (Table 2). In the iTRAQ experiment with cell lines, peptide labeling efficiency was 95% (Supplementary Table 2).

5.1.3.2 Relative quantification of the proteomes of colorectal tissues and cell lines.

The concentration range for proteins expressed in human tissues spans ten orders of magnitude. We chose not to deplete our protein samples of high-abundance proteins (e.g., albumin, IgG), because with the number of tissue samples being analyzed, additional sample preparation steps were considered potential sources of confounding variability (30). As an alternative, each of the ten pooled iTRAQ-labeled samples (ten experiments) were separated into 12 fractions based on the isoelectric point of peptides, reducing the complexity of our protein matrix and limiting the risk of a bias toward the more abundant proteins.

The expression levels of the 4325 non-redundant protein families we were able to relatively quantify in colorectal tissues spanned four orders of magnitude, as deduced from the protein Mascot emPAI value (used as a proxy for the emPAI value (31) to estimate protein concentrations) (Figure 2A).

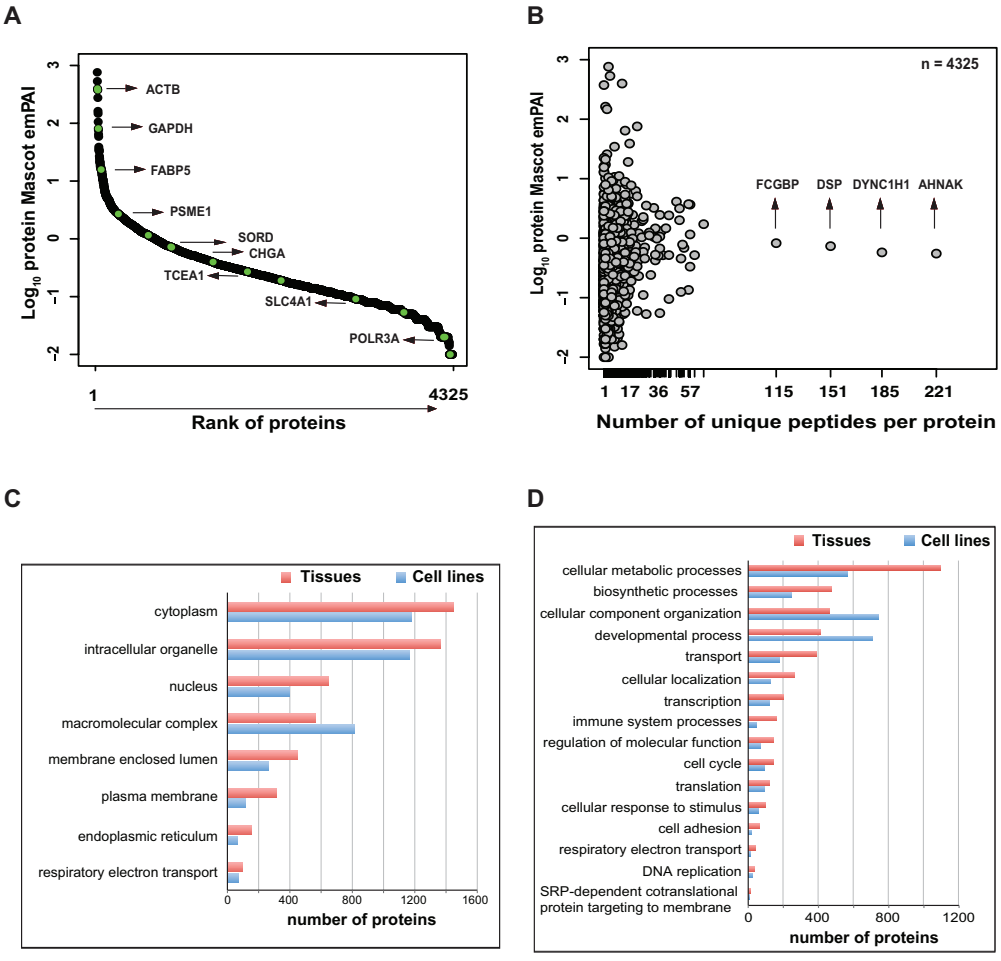


Figure 2. Protein coverage with iTRAQ shotgun analysis in colorectal tissues

(A) Analysis of Mascot emPAI values (used as a proxy for emPAI values) revealed a dynamic range of protein abundance in tissues that spanned four orders of magnitude (y-axis) and corresponded with known abundance estimates for various proteins in these tissues. The high/moderate-abundance proteins (e.g., ACTB, FABP5, CHGA) and low-abundance protein (e.g., POLR3A) relatively quantified in our samples are highlighted relative to their mean Mascot emPAI value. (B) Distribution of reported abundance ranges for the proteins with ≥1 unique peptides identified in tissues, and the high-MW proteins with the highest number of unique peptides identified. (C) Subcellular localizations of the proteins identified in colorectal tissues and cell lines and (D) biological processes in which these proteins are involved. This analysis was performed using Scaffold and Gene Ontology annotations (see Experimental procedures).

Thirty percent (1304/4325) of these families were relatively quantified on the basis of more than one unique peptide. At the top of this list were the large proteins AHNAK, DYNC1H1, DSP, and FCGBP (Figure 2B). In colon epithelial cell lines, 1174 of the 2017 protein families were relatively quantified with more than one unique peptide.

Gene Ontology Annotation in Scaffold was used to identify the subcellular localizations of these protein families and the biological processes they were involved in. The GO categories represented in the tissue and cell line proteomes were fairly similar. In the cell-line proteome, however, the categories generally contained fewer proteins since the total number of proteins detected in these cells was lower than that of the tissues (Figure 2C). Cytoplasmic and organelle- or membrane-associated proteins were the most highly represented categories in our extracts, but nuclear proteins were also readily identified, which indicates that our protein extraction procedure was not strongly biased toward a few cell compartments. The most highly represented biological processes in the tissue proteome were metabolic or biosynthetic processes, whereas cell component organization and developmental processes predominated in the cell line proteome (Figure 2D). Stromal contamination is probably responsible for the increased representation of immune system processes in the tissue proteome (compared with that of the cell lines).

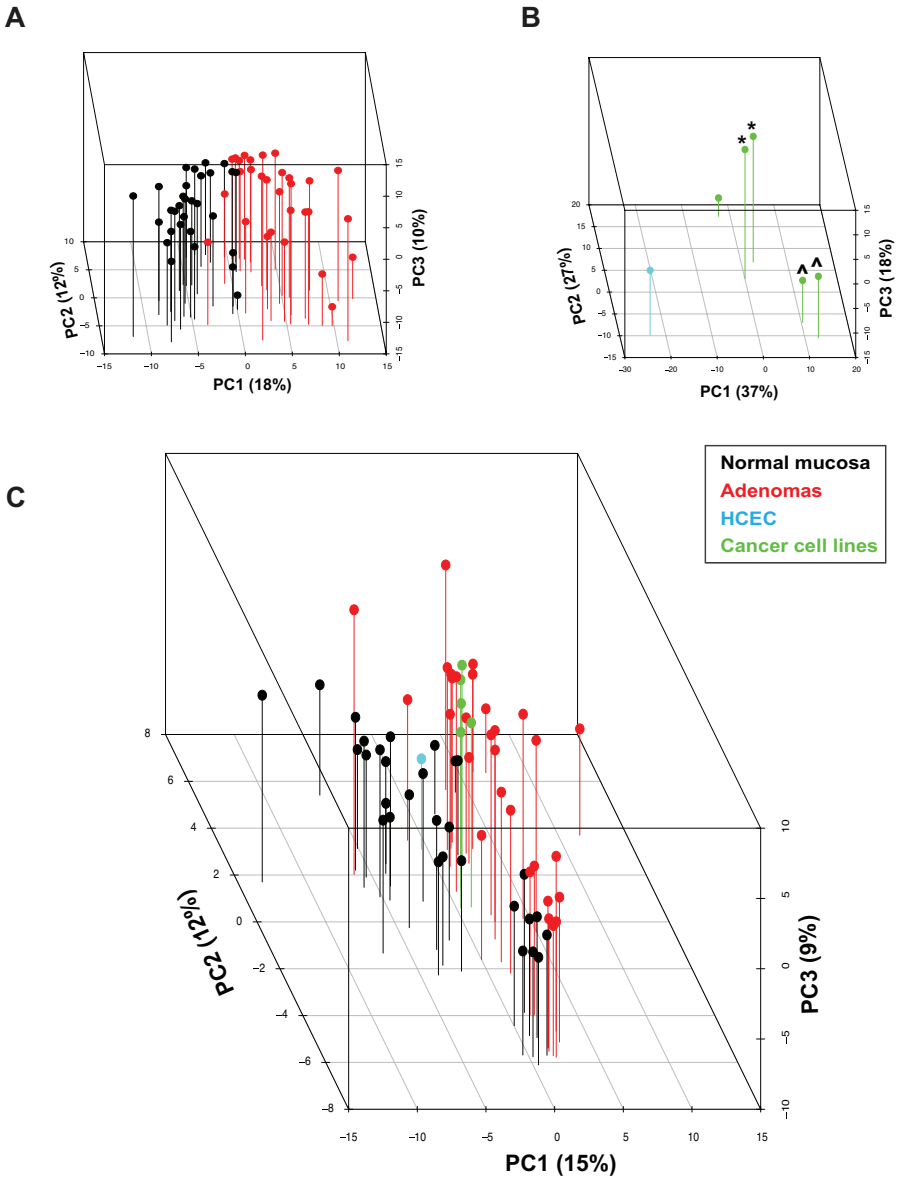


Figure 3. Principal component analysis of protein expression

Three-dimensional principal component analysis (PCA) score plot of log2 protein expression intensity values for (A) tissues (normal mucosa, black; adenomas, red); (B) cell lines (HCEC, cyan; colon cancer cell lines, green); and (C) both. The first three principal components (PCs) account for 40%, 82%, and 36% of the total variance in the tissue, cell line, and tissue+cell line sets, respectively. PC1, the main direction of spread in the three groupings, reflects intergroup variance based on tissue or cell-line type (i.e., normal/immortalized versus tumorous). Cell lines derived from the same patient: * SW480 and SW620 cells; ^ HT29 and CX1 cells.

Log_2-expression levels of the protein families identified in all tissues (n=1072) and cell lines (n=1957) (Table 2) were subjected to principal component analysis (PCA), which easily distinguished the adenomas from the normal mucosa samples (Figure 3A) and the five colon cancer cell lines from the immortalized normal colon epithelial cell line HCEC (Figure 3B). The cancer cell lines were also segregated into three distinct

groups reflecting their patient origins (Figure 3B). When PCA was performed on the expression intensity values of the 1496 non-redundant proteins expressed and quantified in all tissues and cell lines (i.e., those representing the intersection of the tissue [n=10,452] and cell-line [n=5056] protein sets reported in Table 2), colon cancer cell lines clustered with adenomas, while HCEC cells were closer to the normal mucosa samples (Figure 3C).

As a quality control measure, data for the 60 tissue samples (1072 protein families) were subjected to hierarchical clustering analysis. As shown in Supplementary Figure 1, three main clusters emerged: one consisting almost exclusively of normal mucosa samples, a second containing mainly adenomas, and a third that included both tissue types. The 18 samples in the third cluster (nine adenoma/normal mucosa pairs) formed three sub-clusters, which corresponded to three of the 10 experiments for which trypsin digestion, iTRAQ labeling, and LC-MS/MS were performed on the same day. These findings were suggestive of an experimental bias. Indeed, when these 18 potentially sub-standard samples were included in subsequent statistical analyses, they diminished the stringency of our threshold and increased the error margin for false identification. We therefore excluded these samples from the analyses described in the following section.

Proteins displaying dysregulated expression in colorectal adenomas and colon cell lines

To identify proteins whose expression was significantly altered in adenomas (relative to normal mucosa), we analyzed data on the proteins quantified in the remaining 21 tissue pairs. The experimentally derived protein fold-change threshold defining differential expression was based on comparison of the distributions of average intensity $\log_2$ ratios in the reference standard (113 vs. 114, seven experiments) and in patient samples (adenoma vs. normal, seven experiments). The average ratios in the reference sample were centered around 1 (i.e., $\log_2 0$). Average fold-change ratios for the tissue samples displayed wider variance (Supplementary Figure 2). Seventeen percent of the $\log_2$ ratios for the tissue samples exceeded ± 0.5 $\log_2$ scale (indicating a linear fold change of ≥ ± 1.4) as opposed to only 5% of those for the reference samples. For each protein, a paired t-test was used to compare the intensity ratios in normal and adenomatous samples (i.e., normal/114; adenoma/114). After adjustment for multiple comparison (Benjamini-Hochberg method), we selected a stringent q value cutoff of ≤ 0.02.

The 212 proteins that satisfied this criterion and presented a mean expression fold change of ± 1.4 (log_2 0.5) or more were classified as significantly dysregulated in adenomas. They included 76 with upregulated expression and 136 with downregulated expression in the tumor samples (Table 3).

Table 3. Proteins displaying differential expression in adenomas vs. normal mucosa

UniProt accession no.*	Gene name	q value	Average fold change (log2)
P12429	ANXA3	0.00000001	1.44
Q9UN36	NDRG2	0.00000001	-0.79
P00918	CA2	0.00000003	-2.26
P23946	CMA1	0.00000005	-1.38
P00488	F13A1	0.00000005	-1.30
Q9UBR2	CTSZ	0.00000016	-1.23
P10645	CHGA	0.00000016	-1.82
O60844	ZG16	0.00000024	-2.02
P17174	GOT1	0.00000045	-0.61
P31949	S100A11	0.00000062	1.49
P00338 ^; **P07195**	LDHA; LDHB	0.00000068	0.62
O60701	UGDH	0.00000104	-0.70
P55011	SLC12A2	0.00000115	0.85
O95571	ETHE1	0.00000120	-0.74
Q01105	SET	0.00000230	0.72
Q16851	UGP2	0.00000275	-0.53
Q00796	SORD	0.00000275	0.62
P20231; Q15661	TPSB2; TPSAB1	0.00000275	-1.51
Q15181; **Q9H2U2 ^**	PPA1; PPA2	0.00000421	0.64
Q9H3G5	CPVL	0.00000467	-0.69
P01282	VIP	0.00000722	-1.10
P07339	CTSD	0.00000856	-0.71
P19338	NCL	0.00000934	0.50
Q6UWP2	DHRS11	0.00001070	-1.20
P04066	FUCA1	0.00001240	-1.32
Q53EL6	PDCD4	0.00001270	-0.51
P07585	DCN	0.00001580	-1.36
P02511	CRYAB	0.00002720	-1.19
Q96CX2	KCTD12	0.00003590	-0.70
Q05707; **P08123**	COL14A1	0.00004020	-1.21
P51884	LUM	0.00004160	-0.96
Q15063	POSTN	0.00004240	-2.10
P21397	MAOA	0.00004240	-0.63
O00748	CES2	0.00004610	-0.88
Q56VL3	OCIAD2	0.00004960	0.95
Q9BYZ8	REG4	0.00004960	0.98
P55008	AIF1	0.00005650	-0.75
P50224; **P50225**	SULT1A3; SULT1A1	0.00005760	-0.63
O14773	TPP1	0.00006430	-0.52
Q16853	AOC3	0.00006650	-1.16
P53634	CTSC	0.00006940	-0.58
O95881	TXNDC12	0.00006940	0.55
O75795	UGT2B17	0.00006940	-1.57
O00391	QSOX1	0.00006940	-0.70
Q99538	LGMN	0.00006940	-0.63
P12111	COL6A3	0.00006990	-0.74
P80188	LCN2	0.00006990	1.32
P12956 ^	XRCC6	0.00007190	0.53
Q6NZI2	PTRF	0.00007860	-0.76
P09382	LGALS1	0.00007880	-0.84
P25815	S100P	0.00008020	3.38
Q15118	PDK1	0.00008120	-0.90
O75380	NDUFS6	0.00008650	-0.63
Q9HAW8	UGT1A10	0.00009090	-0.95
P01042	KNG1	0.00009660	-0.74
O75356	ENTPD5	0.00012385	-0.54

Table 3. (continued).

Q15293	RCN1	0.00012867	0.61
P17931	LGALS3	0.00013151	-0.62
P36952	SERPINB5	0.00013151	1.31
P09211 ^	GSTP1	0.00014809	0.55
P20774	OGN	0.00015467	-1.52
Q8NBJ4	GOLM1	0.00015467	-0.51
Q16563	SYPL1	0.00015568	1.16
P04229; P13761; Q30134; Q9TQE0; Q9GIY3; Q29974; P04440; P79483	HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DPB1; HLA-DRB3	0.00015568	-0.81
P35555	FBN1	0.00017230	-0.74
Q9H4M9; **Q9NZN4**; **Q9H223**;	EHD1; EHD2; EHD4	0.00017230	-0.56
O00754	MAN2B1	0.00017278	-0.65
B4DR31	DPYSL2	0.00018023	-0.55
P06703 ^	S100A6	0.00018025	1.18
P19801	ABP1	0.00022289	-0.50
P20042	EIF2S2	0.00022417	0.68
P51858	HDGF	0.00024151	0.56
P49959 ^	MRE11A	0.00024174	0.64
O00299 ^; **Q9Y696**	CLIC1; CLIC4	0.00024726	0.67
P61604 ^	HSPE1	0.00024977	0.52
O43252; O95340	PAPSS1; PAPSS2	0.00027464	-0.57
Q12765	SCRN1	0.00029349	-0.64
Q9HB40	SCPEP1	0.00030328	-0.87
P48556	PSMD8	0.00032019	0.56
O60547	GMDS	0.00034711	0.58
Q9Y6R7	FCGBP	0.00035167	-0.73
P61626	LYZ	0.00036558	0.74
Q9Y224	C14orf166	0.00038868	0.52
P01765; P01764	(Ig heavy chain V-III region TIL; Ig heavy chain V-III region VH26)	0.00038868	-0.69
Q9NVP1	DDX18	0.00038996	0.67
P80365	HSD11B2	0.00039413	-0.68
P39687; **Q92688**	ANP32A; ANP32B	0.00039654	0.56
Q86WA6	BPHL	0.00046215	0.52
P24298	GPT	0.00047338	-0.56
Q12874	SF3A3	0.00052166	0.50
P04899; **Q14344**; **P63092**	GNAI2; GNA13; GNAS	0.00062192	-0.51
Q15124	PGM5	0.00068593	-0.68
Q9HAW7; **Q9HAW9**; **O60656**	UGT1A7; UGT1A8; UGT1A9	0.00071202	-0.87
P19224	UGT1A6	0.00071202	-0.87
P06748	NPM1	0.00071508	0.81
Q9NUV9	GIMAP4	0.00071508	-0.89
P18283 **^**	GPX2	0.00071508	0.54
P13688	CEACAM1	0.00072482	-1.06
P01591	IGJ	0.00084082	-1.10
P19823	ITIH2	0.00085565	-0.71
P01774; P01776; P01779	(Ig heavy chain V-III region POM; Ig heavy chain V-III region WAS; Ig heavy chain V-III region TUR)	0.00085565	-0.79
P27695 ^	APEX1	0.00086971	0.54
Q9C002	NMES1	0.00087016	-0.91
Q96F85	CNRIP1	0.00088956	-1.27
Q9BPX5	ARPC5L	0.00095098	0.67
P62263	RPS14	0.00095182	0.52
Q9BY32	ITPA	0.00095634	0.51
P01625	(Ig kappa chain V-IV region Len)	0.00097041	-0.61

Table 3. (continued).

Q15582	TGFBI	0.00097041	0.90
Q07021 ^	C1QBP	0.00101874	0.76
P00738	HP	0.00102030	-0.61
O15143	ARPC1B	0.00103757	-0.50
Q03154	ACY1	0.00108571	0.60
Q9HCB6	SPON1	0.00115609	-0.89
Q96HE7	ERO1L	0.00119027	0.50
P08575	PTPRC	0.00119950	-0.52
Q9Y266	NUDC	0.00152358	0.56
P63313; P62328	TMSB10; TMSB4X	0.00159602	0.98
Q96SQ9	CYP2S1	0.00162405	0.84
Q71U36	TUBA1A	0.00162405	0.54
P00915	CA1	0.00163258	-1.18
P04844	RPN2	0.00165748	0.55
P09669	COX6C	0.00171230	-0.61
P21980	TGM2	0.00174251	-0.55
P00325	ADH1B	0.00175658	-1.18
O14745	SLC9A3R1	0.00175658	-0.51
Q9H8H3	METTL7A	0.00179938	-0.50
P61009	SPCS3	0.00186488	-0.69
Q15746; O15264; Q16539	MYLK; MAPK13; MAPK14	0.00186488	-0.53
P01876; P01877; **Q92973**	IGHA1; IGHA2; TNPO1	0.00191760	-1.01
P12109	COL6A1	0.00194792	-0.67
Q9BX66	SORBS1	0.00205902	-0.64
E9PGJ9	CC2D1A	0.00213982	-0.53
P49006	MARCKSL1	0.00233532	0.51
Q01524	DEFA6	0.00233532	1.60
P01620	(Ig kappa chain V-III region SIE)	0.00238585	-0.56
P36873	PPP1CC	0.00240171	0.58
Q07507	DPT	0.00240576	-1.28
P37840	SNCA	0.00261720	-0.57
P00326; P07327	ADH1C; ADH1A	0.00271753	-1.17
P22105	TNXB	0.00271753	-0.51
O95299	NDUFA10	0.00272901	-0.84
Q9NRP0	OSTC	0.00275806	0.76
P10082	PYY	0.00282902	-1.59
P21810	BGN	0.00283401	-0.66
Q8IV08	PLD3	0.00295758	-0.74
P01857; P01859; P01860; P01861	IGHG1; IGHG2; IGHG3; IGHG4	0.00343621	-0.57
P62330	ARF6	0.00343645	0.84
Q03135	CAV1	0.00346186	-0.71
P22309; P35503; P22310; P35504	UGT1A1; UGT1A3; UGT1A4; UGT1A5	0.00349148	-0.87
Q9NSU2	TREX1	0.00349148	-0.76
Q9UKY7	CDV3	0.00355684	0.79
Q7Z4V5	HDGFRP2	0.00360855	0.65
P07357	C8A	0.00372327	-0.62
Q99757	TXN2	0.00377179	-0.73
P13686	ACP5	0.00385765	-0.80
Q8WWA0	ITLN1	0.00392858	-1.30
P62861	FAU	0.00395014	0.51
P57737	CORO7	0.00401004	-0.83
P10606	COX5B	0.00412793	-0.71
Q9Y259	CHKB	0.00420774	-0.59
Q9Y2J8	PADI2	0.00435864	-0.50
O94919	ENDOD1	0.00451988	-0.80
B9A064; P0CG05; A0M8Q6	IGLL5; IGLC2; IGLC7	0.00451988	-0.56
P20039	HLA-DRB1	0.00459324	-0.82

Table 3. (continued).

P63167 ^; **Q96FJ2**	DYNLL1; DYNLL2	0.00468484	0.63
Q9UNN8	PROCR	0.00480932	-0.78
P07099	EPHX1	0.00486543	-0.54
P32322	PYCR1	0.00495977	0.55
Q9P0J0	NDUFA13	0.00534323	-0.60
E7EUF8; **E9PFN5**	EPB41L3; GSTK1	0.00539222	-0.95
O75531	BANF1	0.00560405	0.73
P26447	S100A4	0.00562754	-0.53
Q9NVJ2	ARL8B	0.00562754	0.50
Q8N752	CSNK1A1	0.00562754	0.52
P40616	ARL1	0.00583778	0.60
Q96GA7	SDSL	0.00583778	-0.82
P01275	GCG	0.00607808	-1.33
P15289	ARSA	0.00633336	-0.57
O75521	ECI2	0.00635218	-0.60
P62158 ^	CALM1; CALM2; CALM3	0.00657472	0.67
P49821	NDUFV1	0.00669319	-0.66
Q15746-5	MYLK	0.00678109	-0.51
Q96BM9	ARL8A	0.00686655	0.54
Q6UX06	OLFM4	0.00696505	1.14
P10153	RNASE2	0.00724902	-0.50
P19075 **^**	TSPAN8	0.00837908	0.59
Q8WU39	PACAP	0.00837978	-0.56
P21953	BCKDHB	0.00837978	0.54
O76041	NEBL	0.00837978	0.71
Q9H4G4	GLIPR2	0.00849532	-1.10
P01766; P01767; P01768	(Ig heavy chain V-III region BRO;Ig heavy chain V-III region BUT; Ig heavy chain V-III region CAM)	0.00895132	-0.61
Q9NR56; Q5VZF2	MBNL1; MBNL2	0.00996176	0.55
P27105	STOM	0.01083127	-0.51
P05387	RPLP2	0.01100903	0.62
Q96AB3	ISOC2	0.01164408	-0.51
O43294	TGFB1I1	0.01198321	-0.57
Q08752	PPID	0.01211602	0.55
Q96DG6	CMBL	0.01289211	-0.51
P61619	SEC61A1	0.01375705	0.59
P56381; **Q5VTU8**	ATP5E; ATP5EP2	0.01440856	-0.52
P14174 ^	MIF	0.01488262	0.51
P12110	COL6A2	0.01526347	-0.53
Q14956	GPNMB	0.01546825	-0.63
P46952	HAAO	0.01570996	-0.53
Q86VN1	VPS36	0.01610077	0.67
Q96S52	PIGS	0.01626862	-0.61
P15559 ^	NQO1	0.01626862	0.56
O60575	SPINK4	0.01810104	0.77
P55735	SEC13	0.01827155	0.59
P02452	COL1A1	0.01933726	-1.32
P00403	MT-CO2	0.02012815	-0.62

* Two or more accession numbers: proteins from the same family or isoforms from the same gene. Boldface numbers indicate "epithelial cell signature" proteins. (See text.)

^ Designated candidate cancer biomarkers in the Human Protein Atlas database

When protein abundance iTRAQ ratios for these 212 proteins were plotted on a heat map, adenomas and normal mucosa samples formed two distinct clusters (Figure 4A).

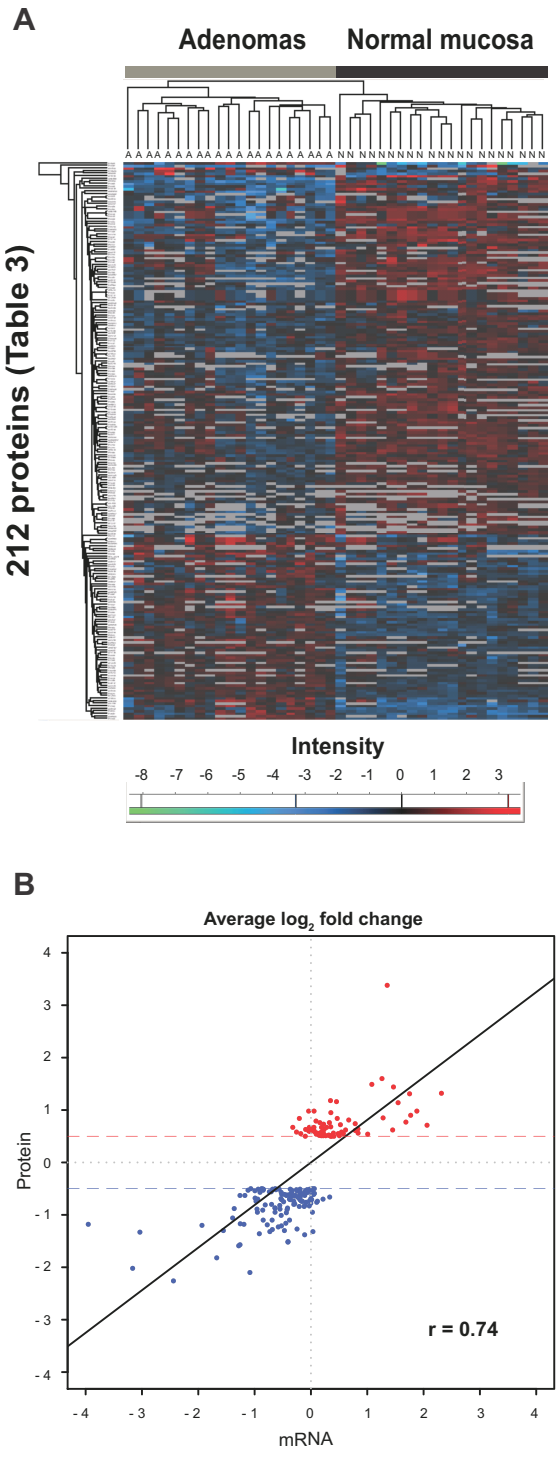


Figure 4. Analysis of the 212 proteins displaying significant tumor-related dysregulation.

(A) Hierarchical clustering of iTRAQ abundance ratios (normal vs. 114, adenoma vs. 114) for the 212 proteins displaying significant adenoma-related dysregulation grouped tissue samples into two discrete clusters: adenoma (A) and normal (N). (B) Pearson's correlation test comparing average fold changes (≥ ±0.5 $\log_2$) for the 212 proteins (red, upregulated; blue, downregulated) in the tissue series with average $\log_2$ fold changes for the corresponding mRNAs measured in another set of adenoma/normal mucosal samples.

As shown in Figure 4B, tissue expression levels for the 212 dysregulated proteins showed good correlation ($r = 0.74$, $P < 0.001$, 95% CI: 0.67 - 0.79) with those of mRNAs for the same genes (measured by our group in another set of colorectal adenomas) (26).

Table 4 lists the biological processes that were over-represented in this set of proteins. At the top of this list was xenobiotic metabolism, a process already linked with adenoma formation on the basis of enrichment studies of transcriptomic datasets conducted by our group (32). Three of the dysregulated proteins involved in this process (CYP2S1, NQO1, and GSTP1) displayed upregulated expression in adenomas, but most were characterized by tumor-related downregulation (ADH1B, ADH1C/ADH1A, UGT1A9/UGT1A6, UGT1A1/UGT1A4/UGT1A3/UGT1A5, UGT1A7/UGT1A8, UGDH, MAOA, SULT1A3/SULT1A1, PAPSS1/PAPSS2, UGP2). Network-building analysis revealed that all these proteins were linked by sub-networks controlled by cancer-associated transcription factors, such as SP1 or, less frequently, MYC, HIF1A, or TP53 (Supplementary Figure 3). As noted in Table 4, a very similar picture emerged when gene ontology enrichment was also analyzed in a larger set of 621 dysregulated proteins selected with less stringent criteria (q value cut-off ≤ 0.2; average $\log_2$ fold change ≥ ± 0.5).

Table 4. Gene ontology (GO) biological processes enriched in the set of 212 proteins whose expression displayed adenoma-related dysregulation (*see Table 3*).

GO ID	GO Term	Annotated *	Significant *	Up in adenomas	Down in adenomas	Expected *	*elim* P value *
GO:0006805	**xenobiotic metabolic process †**	64	21	3	18	2.8	1.70E-09
GO:0006958	**complement activation, classical pathway †**	74	17	0	17	3.24	1.10E-08
GO:0051552	**flavone metabolic process †**	5	5	0	5	0.22	1.50E-07
GO:0052696	**flavonoid glucuronidation †**	5	5	0	5	0.22	1.50E-07
GO:0052697	**xenobiotic glucuronidation †**	5	5	0	5	0.22	1.50E-07
GO:0042573	**retinoic acid metabolic process †**	9	6	0	6	0.39	5.00E-07
GO:0045087	**innate immune response §**	249	33	4	29	10.91	7.80E-07
GO:0031295	**T cell costimulation §**	34	10	0	10	1.49	1.10E-06
GO:0030199	**collagen fibril organization †**	8	5	0	5	0.35	7.80E-06
GO:0001501	**skeletal system development**	54	11	1	10	2.37	1.60E-05
GO:0070208	**protein heterotrimerization †**	5	4	0	4	0.22	1.70E-05
GO:0050852	**T cell receptor signaling pathway §**	48	10	0	10	2.1	3.20E-05

*** Annotated**: proteins in TopGO Background list; **Significant**: 212 dysregulated proteins of Table 3; **Expected**: Number of signifcant proteins expected to map to the GO term if the significant proteins were randomly distributed over all GO terms. ***elim* P value**: P value from the elim method (ref. no. 25). Only processes with an *elim* P value < 1.0E-04 are shown.

† Processes that were also among the top 12 processes displaying enrichment in a larger set of 621 dysregulated proteins selected with less stringent criteria (q value ≤ 0.2; average $\log_2$ fold change ≥ ± 0.5; see *Results* section for details); § processes that shared a common GO ancestor (immune system process) with the process displaying most significant enrichment in the larger set.

The expression levels of 111 (52%) of the 212 proteins that were differentially expressed in adenomas were also quantified in cell lines (those shown in boldface in Table 3 and referred to hereafter as the "epithelial cell signature" proteins). Almost half (n=51, 46%) showed directionally similar tumor-related dysregulation in both analyses. Since cell-line studies were conducted with only one non-cancerous line, these findings obviously require further validation. They suggest, however, that these 51 proteins are indeed expressed in the epithelial cells of normal colorectal tissues and that their expression is dysregulated in the epithelial cells of adenomas.

5.1.3.3 Upregulation of SORD expression and activity in colorectal adenomas and cancer cell lines

Sorbitol dehydrogenase, a key enzyme in the polyol pathway, was one of the most significantly upregulated proteins in our colorectal adenomas (based on q values) (Table 3). Because its increased expression could have metabolic consequences with

potential impact on tumorigenesis, we performed Western blotting and immunostaining studies to validate this finding. The reliability of the anti-SORD antibody we had chosen was first tested on protein extracts from the six colorectal epithelial cell lines (Figure 5A).

The tumor-related log_2 fold changes detected with Western blotting were substantially larger than those documented with iTRAQ (2 to 6 vs. 0.4 to 1, respectively) (Figure 5B), which was not surprising since iTRAQ has been reported to underestimate protein abundance (33). However, the relative quantities of SORD found with the two methods were fully consistent. As for the 21 adenomas, the elevated SORD expression documented in these tumors by iTRAQ (Figure 5C) showed good correlation with the increased SORD mRNA levels we had previously found in other 42 lesions of this type (26) (Figure 5D). Western blot analysis of four randomly selected adenoma/normal mucosa pairs from the present series revealed obvious upregulation of SORD expression in all four tumors although the magnitude of the increase varied (Figure 5E).

SORD activity was then assayed (see Experimental procedures) to see how it corresponded with the enzyme expression levels reported above. As shown in Figure 5F, the results of cell line assays were fully consistent with the Western blotting data: SORD activity was 7 times higher in HT29 than in HCEC cells, and more limited upregulation was found in SW480. High correlation between enzyme activity and protein level was also documented for three randomly selected adenoma/normal mucosa pairs (Figure 5F).

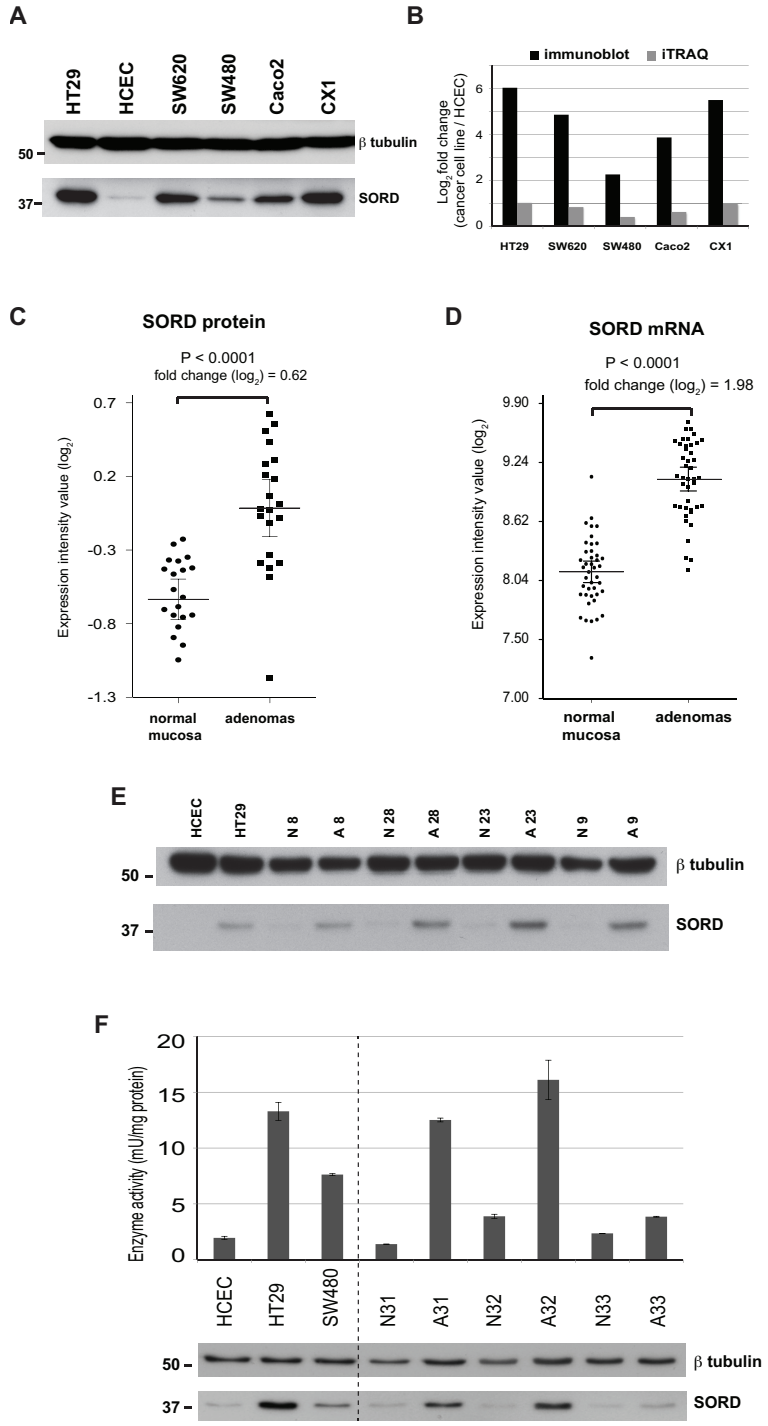


Figure 5. Significantly upregulated SORD expression and activity in colorectal cell lines and adenomas.

(A) Tumor-related upregulation of sorbitol dehydrogenase (SORD) in colon cancer cell lines was confirmed with Western blotting. The SORD dysregulation trend was identical to that observed with iTRAQ-based MS/MS, although when immunoblot results were quantified (B), the $\log_2$ fold changes were over five times greater than those documented in the iTRAQ study. (C) SORD protein expression (iTRAQ analysis) in 21 normal mucosa/adenoma tissue pairs. (D) SORD mRNA expression in 42 other normal mucosa/adenomas from a previous study by our group (26). Error bars indicate the means and 95% confidence intervals. (E) Western blots showing tumor-related upregulation of SORD expression in

four randomly selected adenoma [A] /normal mucosa [N] tissue pairs of the 21 shown in panel C (see Table 1 for sample descriptions). (F) SORD activity also displayed tumor-related upregulation in cell lines (HT29 and SW480 versus HCEC cells) and tissues (adenomas versus normal mucosa). Columns show mean enzyme activity measured in at least two replicates; error bars indicate standard deviations from means. The Western blot beneath the graph shows SORD levels measured in the extracts used for the enzyme activity assays.

MS and Western blotting findings were further validated with immunostaining studies, as shown in Figure 6. Cytoplasmic SORD staining was evident in the colon cancer cell line HT29 but weaker or even absent in normal epithelial HCEC cells (Figures 6A and 6B). As for colorectal tissues, SORD cytoplasmic expression was limited to the bottom of the normal epithelial crypts (Figures 6C, 6D, and 6E), but its expression was markedly increased in adenomatous and cancerous glands (Figures 6F - 6I). These findings suggests that SORD is likely to be expressed in proliferating cells, although it was largely absent in HCECs, which undergo regular proliferation in vitro. Furthermore, nuclear localization of SORD was noted in some adenomatous crypts (Supplementary Figure 4A and C), and the cells in question were almost always negative for the well-known proliferation marker Ki-67 (Supplementary Figure 4B and D). This mutually-exclusive staining pattern was also observed in normal crypts of the ileum where SORD, interestingly, appeared to be expressed in the nuclei of putative stem cells (Supplementary Figure 4E and F).

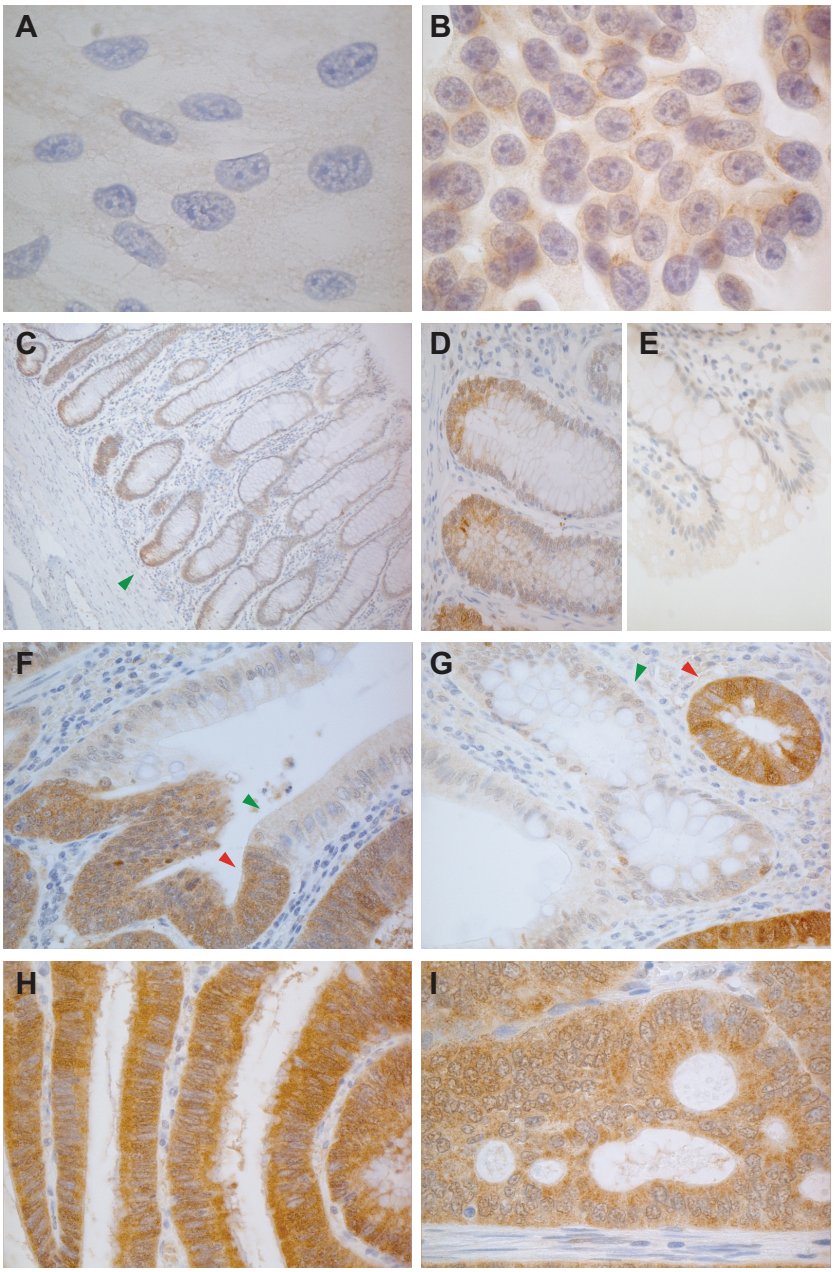


Figure 6. Anti-SORD immunostaining of colorectal cell lines and tissues.

Consistent with proteomic data, SORD expression was (A) negligible or absent in HCECs, but (B) clearly expressed in the cytoplasm of HT29 cells. (C) In normal colorectal mucosa, SORD expression was limited to the lower portion of the epithelial crypts, where stem cells and highly-proliferating cells are located. Higher magnification views show staining at (D) the base vs. (E) mouth of colonic crypts. (F) and (G): Its expression was markedly increased in adenomatous glands (red arrowheads) compared with normal crypts (green arrowheads). Panels (H) and (I) show abundant expression of SORD in a large adenoma and in a cancer, respectively.

5.1.3.4 Polyol pathway enzyme expression and metabolite levels in cell lines and tissues.

We then examined the state of the polyol pathway (Supplementary Figure 5A) in colorectal cell lines and tissues. As shown in Supplementary Figure 5B, immunoblot

studies revealed decreased AKR1B1 expression in HT29 (vs. HCEC cells) and adenomas (vs. corresponding normal mucosal samples), while SORD expression and that of KHK were upregulated in tumor cells and tissues. As for the metabolites (Supplementary Figure 5C), D-glucose levels were significantly decreased in adenomas. Less dramatic changes were observed in the levels of D-sorbitol and D-fructose, which both showed a tendency to decrease in tumor tissues.

5.1.4 DISCUSSION

Although a number of proteomic studies have comparatively analyzed different types of colorectal tissues, precancerous lesions have been considered in only three (21, 34, 35), and in two of these (21, 34), the number of adenomas analyzed was very small (≤ 4). The study by Lam et al. (35) is the only one that compared protein expression in a relatively large number (n=20) of paired adenoma and normal mucosa samples. They used 2-DE to resolve over 1000 proteins in the two tissue groups, and those displaying differential expression were then analyzed with MALDI-TOF/TOF MS. MS/MS validation pinpointed four proteins (ANXA3, S100A11, EIF-5A1, S100P) whose expression in adenomas was significantly increased. Using MS with iTRAQ 8-plex peptide labeling and OFFGEL fractionation allowed us to quantitatively compare protein expression in 30 colorectal adenomas and paired samples of normal mucosa, and to investigate low-abundance proteins that cannot be evaluated with 2-DE-based proteomics. All in all, 4325 non-redundant protein families were quantified in our colorectal tissues (25% of which were identified in all 60 samples) (Table 2). And the 212 proteins we flagged as being significantly dysregulated in adenomas included three of the four proteins identified by Lam et al. (upregulation of the fourth, EIF-5A1, failed to meet our stringent criterion for significance) (Table 3).

The cell types in which these proteomic changes occur is of obvious interest since colorectal cancer arises from the epithelial component of the colorectal mucosa. Although our findings are preliminary and will naturally require validation in future studies, 51 of the 212 proteins listed in Table 3 were "epithelial cell signature" proteins and showed directionally similar expression changes in colon cancer cell lines vs. HCEC. It therefore seems likely that their dysregulated expression in adenomas is a feature of neoplastic transformation of colorectal epithelial cells. However, epithelial-stromal cell interactions can also play important roles in tumorigenesis (20). Our approach also allowed us to identify 101 proteins displaying adenoma-related dysregulation that were probably of stromal-cell origin since they were not expressed in any of the six epithelial cell lines we examined (Table 3). These proteins were mainly involved in immune-related processes (immune response, complement activation, T cell co-stimulation), which are usually not represented in colon epithelial cell lines. Their expression changes are likely to have important effects on the microenvironment of an epithelial-cell tumor.

Our search for potential biomarkers of early-stage colorectal tumorigenesis focused exclusively on the 76 proteins whose expression was significantly upregulated

in adenomas. According to the Human Protein Atlas (HPA) database (36), 69 (91%) of these have cancer-related features, and 16 of the 69 are already classified as candidate cancer biomarkers (Table 3). The HPA database contains information on protein expression in normal and cancer tissues, but not in those regarded as precancerous. The overlap between our findings and those of the HPA suggest that most protein expression changes identified thus far in colorectal adenocarcinomas are probably already detectable in the benign precursors of these lesions.

Supplementary Figure 6 shows the expression profiles of the 10 proteins that were most markedly upregulated in adenomas. This group comprised two of the four proteins identified by Lam et al. (35) as significantly overexpressed in adenomas. **Annexin A3 (ANXA3),** for example, is at the top of our list (based on q values) (Table 3). An angiogenic factor that induces VEGF production via the HIF-1 pathway (37), ANXA3 belongs to a family of calcium-dependent, phospholipid-binding proteins involved in diverse biological processes, including signal transduction, inflammatory responses, membrane organization, and the regulation of cellular growth (38, 39). Dysregulated ANXA expression is also a common feature of colorectal cancer (39) and most other cancers as well (40). **S100A11** expression was also increased in these tumors, which is consistent with earlier reports (41). The cytosolic S100 proteins interact directly with peptides on the N-terminal domain of annexins (38, 42), and like the annexins, they also have diverse intracellular and extracellular functional roles (43).

Among the other top ten proteins displaying adenoma-related upregulation is **LDHA / LDHB**. Their expression levels were not measured separately, but LDHA is presumably responsible for the increased expression observed in our adenomas. LDHB expression is in most cases epigenetically silenced in colon cancer cells (44, 45), whereas LDHA is over-expressed, and its activity is maintained via the oncogenic tyrosine kinase FGFR1 (46). LDHA is a key player in the reversible conversion of pyruvate to lactate during aerobic glycolysis, a typical feature of cancer cell metabolism first described by Otto Warburg (47). The sodium- and potassium-coupled chloride cotransporter, **SLC12A2**, is expressed on the basolateral membrane of the normal colon epithelium, where its recruitment and activation are regulated by calcium and cAMP. Loss of SLC12A2 leads to impaired chloride secretion in the intestine (48, 49), but to our knowledge, there are no published data linking this protein to colon cancer. The fifth markedly over-expressed protein is **SET**, one of the five proteins that make up the inhibitor of acetyltranferases (INHAT) complex. Two other INHAT components, **APEX1** and **ANP32A/ANP32B**, were also upregulated in adenomas (albeit to a lesser

extent than SET) (Table 3). These changes are noteworthy because INHAT binds directly to histones, preventing their acetylation by histone acetyl transferases (50-52), and loss of histone acetylation is a crucial step in gene silencing (53, 54). Thus far, INHAT's role in cancer has not been widely investigated, but overexpression of the complex components has been observed in serous epithelial ovarian cancer (55). The upregulated expression of **PPA1/PPA2** in our adenomas might play various roles in colorectal tumorigenesis since these proteins are key players in the synthesis of fatty acids, nucleotides, amino acids, and other essential molecules (56). The phosphoprotein nucleolin (**NCL**), an essential protein for proliferating cells (57), appears to regulate several steps in the biogenesis of ribosomes, including transcription, ribosome assembly, and the processing of precursor ribosomal RNA (58-60), all of which might be instrumental to adenoma growth. As for **OCIAD2**, strong immunoreactivity for this protein has been reported in early-stage adenocarcinomas of the lung and in ovarian cancers (61-63), but there are no published data linking it to colorectal tumorigenesis. In contrast, the secreted protein **REG4**, which promotes mitosis and enhances the motility and invasiveness of colon cancer cells, is strongly expressed in these cells and in the serum of patients with colorectal cancer (64-66).

The final protein characterized by marked adenoma-related upregulation was **SORD**, a key enzyme in the polyol metabolic pathway. It was selected for validation studies, because although aberrant polyol pathway activity has been implicated in diabetic complications (67-70) and myocardial ischemia (71), the role of SORD in tumorigenesis was completely unknown. During the execution of this study, however, upregulated SORD expression was reported in prostate cancer (72) and in colorectal adenomas (21), and these findings strengthened our resolve to characterize this phenomenon in colorectal tumorigenesis.

Upregulated SORD expression and activity in adenomas (Figure 5) would enhance the production of fructose (see schematic of Supplementary Figure 5A), thereby increasing the generation of triose sugars and diacylglycerol (intermediates in the glycolytic and lipid signaling pathways, respectively). Fructose is also several times more effective than glucose in promoting intracellular non-enzymatic glycation (73-75), and advanced glycation end products may contribute to the vascular complications of diabetes and other pathologic conditions (67, 76-78). Whether these fructose-driven metabolic events play a role in the development of adenomas is unclear, but the polyol pathway was very active in the adenomatous cells we examined. This activity was also reflected by the concomitant increase of the expression of KHK (Supplementary Figure

5B), the enzyme that catalyzes the transformation of fructose to fructose-1-P, downstream from the polyol pathway.

The effects of these enzymatic changes on sorbitol and fructose concentrations in adenomas need to be investigated in larger tissue series, but our preliminary data suggest that the levels of both are slightly decreased in these lesions (Supplementary Figure 5C). In contrast, our adenomas exhibited dramatically reduced concentrations of glucose, the initial substrate in the polyol pathway (Supplementary Figure 5C). Adenoma-related dysregulation was also noted in the expression of AKR1B1, the enzyme that converts glucose to sorbitol (Supplementary Figure 5B). Exploitation of the polyol pathway to divert carbon from glucose to other energy intermediates might provide adenomatous cells with a selective advantage over normal cells. This pathway might prove to be another means of tumor-related glucose consumption in addition to the well-known glycolytic and pentose phosphate pathways (Supplementary Figure 5A). Advanced cancer cells consume glucose at a much higher rate than normal cells, and much of their energy is generated by aerobic glycolysis rather than by oxidative phosphorylation of glucose in the mitochondria (i.e., the Warburg effect) (79). The predominantly glycolytic phenotype of cancer cells results in low glucose levels and high concentrations of lactate (47, 80, 81). The relative concentrations of lactate in the three adenomas we tested were significantly higher than those found in matched samples of normal mucosa (Supplementary Figure 5C), indicating that the Warburg effect is already evident in precancerous colorectal lesions. Studies involving metabolic flux analysis to monitor the fate of isotopic tracers in in vitro and in vivo systems would provide further insight into the biological roles of the polyol pathway in tumorigenesis.

5.1.5 REFERENCES

1. Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011) Global cancer statistics. CA: a cancer journal for clinicians 61, 69-90

2. Siegel, R., Naishadham, D., and Jemal, A. (2013) Cancer statistics, 2013. CA: a cancer journal for clinicians 63, 11-30

3. Shinya, H., and Wolff, W. I. (1979) Morphology, anatomic distribution and cancer potential of colonic polyps. Ann. Surg. 190, 679-683

4. Cattaneo, E., Baudis, M., Buffoli, F., Bianco, M. A., Zorzi, F., and Marra, G. (2011) Pathways and Crossroads to Colorectal Cancer In: Fitzgerald, R. C., ed. Pre-Invasive Disease: Pathogenesis and Clinical Management, pp. 369-394, Springer, New York

5. Force, U. S. P. S. T. (2002) Screening for colorectal cancer: recommendation and rationale. Ann. Intern. Med. 137, 129-131

6. Kahi, C. J., Rex, D. K., and Imperiale, T. F. (2008) Screening, surveillance, and primary prevention for colorectal cancer: a review of the recent literature. Gastroenterology 135, 380-399

7. Levin, B., Lieberman, D. A., McFarland, B., Andrews, K. S., Brooks, D., Bond, J., Dash, C., Giardiello, F. M., Glick, S., Johnson, D., Johnson, C. D., Levin, T. R., Pickhardt, P. J., Rex, D. K., Smith, R. A., Thorson, A., Winawer, S. J., American Cancer Society Colorectal Cancer Advisory, G., Force, U. S. M.-S. T., and American College of Radiology Colon Cancer, C. (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology 134, 1570-1595

8. Sillars-Hardebol, A. H., Carvalho, B., van Engeland, M., Fijneman, R. J., and Meijer, G. A. (2012) The adenoma hunt in colorectal cancer screening: defining the target. J. Pathol. 226, 1-6

9. de Wit, M., Fijneman, R. J., Verheul, H. M., Meijer, G. A., and Jimenez, C. R. (2013) Proteomics in colorectal cancer translational research: Biomarker discovery for clinical applications. Clin. Biochem. 46, 466-479

10. Patterson, S. D., and Aebersold, R. H. (2003) Proteomics: the first decade and beyond. Nat. Genet. 33 Suppl, 311-323

11. Horth, P., Miller, C. A., Preckel, T., and Wenz, C. (2006) Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. Mol. Cell. Proteomics 5, 1968-1974

12. Wu, W. W., Wang, G., Baek, S. J., and Shen, R. F. (2006) Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. J. Proteome Res. 5, 651-658

13. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422, 198-207

14. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics. Chem. Rev. 101, 269-295

15. Bensimon, A., Heck, A. J., and Aebersold, R. (2012) Mass spectrometry-based proteomics and network biology. Annu. Rev. Biochem. 81, 379-405

16. Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Ana.l Bioanal. Chem. 404, 939-965

17. Wiese, S., Reidegeld, K. A., Meyer, H. E., and Warscheid, B. (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 7, 340-350

18. Gan, C. S., Chong, P. K., Pham, T. K., and Wright, P. C. (2007) Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). J. Proteome Res. 6, 821-827

19. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3, 1154-1169

20. Jimenez, C. R., Knol, J. C., Meijer, G. A., and Fijneman, R. J. (2010) Proteomics of colorectal cancer: overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. J. Proteomics 73, 1873-1895

21. Besson, D., Pavageau, A. H., Valo, I., Bourreau, A., Belanger, A., Eymerit-Morin, C., Mouliere, A., Chassevent, A., Boisdron-Celle, M., Morel, A., Solassol, J., Campone, M., Gamelin, E., Barre, B., Coqueret, O., and Guette, C. (2011) A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. Mol. Cell. Proteomics 10, M111 009712

22. Jankova, L., Chan, C., Fung, C. L., Song, X., Kwun, S. Y., Cowley, M. J., Kaplan, W., Dent, O. F., Bokey, E. L., Chapuis, P. H., Baker, M. S., Robertson, G. R., Clarke, S. J., and Molloy, M. P. (2011) Proteomic comparison of colorectal tumours and non-neoplastic mucosa from paired patient samples using iTRAQ mass spectrometry. Mol. BioSyst. 7, 2997-3005

23. Roig, A. I., Eskiocak, U., Hight, S. K., Kim, S. B., Delgado, O., Souza, R. F., Spechler, S. J., Wright, W. E., and Shay, J. W. (2010) Immortalized epithelial cells derived from human colon biopsies express stem cell markers and differentiate in vitro. Gastroenterology 138, 1012-1021

24. Kall, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Posterior error probabilities and false discovery rates: two sides of the same coin. J. Proteome Res. 7, 40-44

25. Alexa, A., Rahnenfuhrer, J., and Lengauer, T. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22, 1600-1607

26. Cattaneo, E., Laczko, E., Buffoli, F., Zorzi, F., Bianco, M. A., Menigatti, M., Bartosova, Z., Haider, R., Helmchen, B., Sabates-Bellver, J., Tiwari, A., Jiricny, J., and Marra, G. (2011) Preinvasive colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid vs. nonpolypoid). EMBO Mol Med 3, 334-347

27. Truninger, K., Menigatti, M., Luz, J., Russell, A., Haider, R., Gebbers, J. O., Bannwart, F., Yurtsever, H., Neuweiler, J., Riehle, H. M., Cattaruzza, M. S., Heinimann, K., Schär, P., Jiricny, J., and Marra, G. (2005) Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 128, 1160-1171

28. Gerlach, U., and Hilby, W. (1974) Methods of Enzymatic Analysis, 2nd ed. Ed., Academic Press Inc., New York, NY

29. Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A. R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat. Protoc. 1, 387-396

30. Afkarian, M., Bhasin, M., Dillon, S. T., Guerrero, M. C., Nelson, R. G., Knowler, W. C., Thadhani, R., and Libermann, T. A. (2010) Optimizing a proteomics platform for urine biomarker discovery. Mol. Cell. Proteomics 9, 2195-2204

31. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics 4, 1265-1272

32. Maglietta, R., Liuzzi, V. C., Cattaneo, E., Laczko, E., Piepoli, A., Panza, A., Carella, M., Palumbo, O., Staiano, T., Buffoli, F., Andriulli, A., Marra, G., and Ancona, N. (2012) Molecular pathways undergoing dramatic transcriptomic changes during tumor development in the human colon. BMC Cancer 12, 608

33. Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. (2009) iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". J. Proteome Res. 8, 5347-5355

34. Albrethsen, J., Knol, J. C., Piersma, S. R., Pham, T. V., de Wit, M., Mongera, S., Carvalho, B., Verheul, H. M., Fijneman, R. J., Meijer, G. A., and Jimenez, C. R. (2010) Subnuclear proteomics in colorectal cancer: identification of proteins enriched in the nuclear matrix fraction and regulation in adenoma to carcinoma progression. Mol. Cell. Proteomics 9, 988-1005

35. Lam, F., Jankova, L., Dent, O. F., Molloy, M. P., Kwun, S. Y., Clarke, C., Chapuis, P., Robertson, G., Beale, P., Clarke, S., Bokey, E. L., and Chan, C. (2010) Identification of distinctive protein expression patterns in colorectal adenoma. Proteomics: Clin. Appl. 4, 60-70

36. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Bjorling, L., and Ponten, F. (2010) Towards a knowledge-based Human Protein Atlas. Nat. Biotechnol. 28, 1248-1250

37. Park, J. E., Lee, D. H., Lee, J. A., Park, S. G., Kim, N. S., Park, B. C., and Cho, S. (2005) Annexin A3 is a potential angiogenic mediator. Biochem. Biophys. Res. Commun. 337, 1283-1287

38. Gerke, V., Creutz, C. E., and Moss, S. E. (2005) Annexins: linking Ca2+ signalling to membrane dynamics. Nat. Rev. Mol. Cell Biol. 6, 449-461

39. Duncan, R., Carpenter, B., Main, L. C., Telfer, C., and Murray, G. I. (2008) Characterisation and protein expression profiling of annexins in colorectal cancer. Br. J. Cancer 98, 426-433

40. Mussunoor, S., and Murray, G. I. (2008) The role of annexins in tumour development and progression. J. Pathol. 216, 131-140

41. Melle, C., Ernst, G., Schimmel, B., Bleul, A., Mothes, H., Kaufmann, R., Settmacher, U., and Von Eggeling, F. (2006) Different expression of calgizzarin (S100A11) in normal colonic epithelium, adenoma and colorectal carcinoma. Int. J. Oncol. 28, 195-200

42. Lewit-Bentley, A., Rety, S., Sopkova-de Oliveira Santos, J., and Gerke, V. (2000) S100-annexin complexes: some insights from structural studies. Cell Biol. Int. 24, 799-802

43. Donato, R. (2001) S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. Int. J Biochem. Cell Biol. 33, 637-668

44. Thangaraju, M., Carswell, K. N., Prasad, P. D., and Ganapathy, V. (2009) Colon cancer cells maintain low levels of pyruvate to avoid cell death caused by inhibition of HDAC1/HDAC3. Biochem. J 417, 379-389

45. Thorn, C. C., Freeman, T. C., Scott, N., Guillou, P. J., and Jayne, D. G. (2009) Laser microdissection expression profiling of marginal edges of colorectal tumours reveals evidence of increased lactate metabolism in the aggressive phenotype. Gut 58, 404-412

46. Fan, J., Hitosugi, T., Chung, T. W., Xie, J., Ge, Q., Gu, T. L., Polakiewicz, R. D., Chen, G. Z., Boggon, T. J., Lonial, S., Khuri, F. R., Kang, S., and Chen, J. (2011) Tyrosine phosphorylation of lactate dehydrogenase A is important for NADH/NAD(+) redox homeostasis in cancer cells. Mol. Cell. Biol. 31, 4938-4950

47. Koppenol, W. H., Bounds, P. L., and Dang, C. V. (2011) Otto Warburg's contributions to current concepts of cancer metabolism. Nat. Rev. Cancer 11, 325-337

48. Reynolds, A., Parris, A., Evans, L. A., Lindqvist, S., Sharp, P., Lewis, M., Tighe, R., and Williams, M. R. (2007) Dynamic and differential regulation of NKCC1 by calcium and cAMP in the native human colonic epithelium. J. Physiol. 582, 507-524

49. Flagella, M., Clarke, L. L., Miller, M. L., Erway, L. C., Giannella, R. A., Andringa, A., Gawenis, L. R., Kramer, J., Duffy, J. J., Doetschman, T., Lorenz, J. N., Yamoah, E. N., Cardell, E. L., and Shull, G. E. (1999) Mice lacking the basolateral Na-K-2Cl cotransporter have impaired epithelial chloride secretion and are profoundly deaf. J. Biol. Chem. 274, 26946-26955

50. Schneider, R., Bannister, A. J., Weise, C., and Kouzarides, T. (2004) Direct binding of INHAT to H3 tails disrupted by modifications. J. Biol. Chem. 279, 23859-23862

51. Kutney, S. N., Hong, R., Macfarlan, T., and Chakravarti, D. (2004) A signaling role of histone-binding proteins and INHAT subunits pp32 and Set/TAF-Ibeta in integrating chromatin hypoacetylation and transcriptional repression. J. Biol. Chem. 279, 30850-30855

52. Seo, S. B., McNamara, P., Heo, S., Turner, A., Lane, W. S., and Chakravarti, D. (2001) Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. Cell 104, 119-130

53. Kondo, Y., and Issa, J. P. (2004) Epigenetic changes in colorectal cancer. Cancer Metastasis Rev. 23, 29-39

54. Kondo, Y., Shen, L., and Issa, J. P. (2003) Critical role of histone methylation in tumor suppressor gene silencing in colorectal cancer. Mol. Cell. Biol. 23, 206-215

55. Ouellet, V., Le Page, C., Guyot, M. C., Lussier, C., Tonin, P. N., Provencher, D. M., and Mes-Masson, A. M. (2006) SET complex in serous epithelial ovarian cancer. Int. J. Cancer 119, 2119-2126

56. Takahashi, K., Inuzuka, M., and Ingi, T. (2004) Cellular signaling mediated by calphoglin-induced activation of IPP and PGM. Biochem. Biophys. Res. Commun. 325, 203-214

57. Lapeyre, B., Bourbon, H., and Amalric, F. (1987) Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. Proc. Natl. Acad. Sci. U.S.A. 84, 1472-1476

58. Ginisty, H., Amalric, F., and Bouvet, P. (1998) Nucleolin functions in the first step of ribosomal RNA processing. The EMBO journal 17, 1476-1486

59. Ginisty, H., Sicard, H., Roger, B., and Bouvet, P. (1999) Structure and functions of nucleolin. J. Cell Sci. 112 (Pt 6), 761-772

60. Tajrishi, M. M., Tuteja, R., and Tuteja, N. (2011) Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus. Commun. Integr. Biol. 4, 267-275

61. Chien, J., Fan, J. B., Bell, D. A., April, C., Klotzle, B., Ota, T., Lingle, W. L., Gonzalez Bosquet, J., Shridhar, V., and Hartmann, L. C. (2009) Analysis of gene expression in stage I serous tumors identifies critical pathways altered in ovarian cancer. Gynecol. Oncol. 114, 3-11

62. Ishiyama, T., Kano, J., Anami, Y., Onuki, T., Iijima, T., Morisita, Y., Yokota, J., and Noguchi, M. (2007) OCIA domain containing 2 is highly expressed in adenocarcinoma mixed subtype with bronchioloalveolar carcinoma component and is associated with better prognosis. Cancer Sci. 98, 50-57

63. Nagata, C., Kobayashi, H., Sakata, A., Satomi, K., Minami, Y., Morishita, Y., Ohara, R., Yoshikawa, H., Arai, Y., Nishida, M., and Noguchi, M. (2012) Increased expression of OCIA domain containing 2 during stepwise progression of ovarian mucinous tumor. Pathol. Int. 62, 471-476

64. Oue, N., Kuniyasu, H., Noguchi, T., Sentani, K., Ito, M., Tanaka, S., Setoyama, T., Sakakura, C., Natsugoe, S., and Yasui, W. (2007) Serum concentration of Reg IV in patients with colorectal cancer: overexpression and high serum levels of Reg IV are associated with liver metastasis. Oncology 72, 371-380

65. Rafa, L., Dessein, A. F., Devisme, L., Buob, D., Truant, S., Porchet, N., Huet, G., Buisine, M. P., and Lesuffleur, T. (2010) REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. Int. J. Oncol. 36, 689-698

66. Violette, S., Festor, E., Pandrea-Vasile, I., Mitchell, V., Adida, C., Dussaulx, E., Lacorte, J. M., Chambaz, J., Lacasa, M., and Lesuffleur, T. (2003) Reg IV, a new member of the regenerating gene family, is overexpressed in colorectal carcinomas. Int. J. Cancer 103, 185-193

67. Obrosova, I. G. (2005) Increased sorbitol pathway activity generates oxidative stress in tissue sites for diabetic complications. Antioxid. Redox Signaling 7, 1543-1552

68. Ola, M. S., Berkich, D. A., Xu, Y., King, M. T., Gardner, T. W., Simpson, I., and LaNoue, K. F. (2006) Analysis of glucose metabolism in diabetic rat retinas. Am. J. Physiol.: Endocrinol. Metab. 290, E1057-1067

69. Van den Enden, M. K., Nyengaard, J. R., Ostrow, E., Burgan, J. H., and Williamson, J. R. (1995) Elevated glucose levels increase retinal glycolysis and sorbitol pathway metabolism. Implications for diabetic retinopathy. Invest. Ophthalmol. Visual Sci. 36, 1675-1685

70. Gabbay, K. H., Merola, L. O., and Field, R. A. (1966) Sorbitol pathway: presence in nerve and cord with substrate accumulation in diabetes. Science 151, 209-210

71. Hwang, Y. C., Bakr, S., Ellery, C. A., Oates, P. J., and Ramasamy, R. (2003) Sorbitol dehydrogenase: a novel target for adjunctive protection of ischemic myocardium. FASEB journal 17, 2331-2333

72. Szabo, Z., Hamalainen, J., Loikkanen, I., Moilanen, A. M., Hirvikoski, P., Vaisanen, T., Paavonen, T. K., and Vaarala, M. H. (2010) Sorbitol dehydrogenase expression is regulated by androgens in the human prostate. Oncol. Rep. 23, 1233-1239

73. Obrosova, I. G. (2005) Increased Sorbitol Pathway Activity Generates Oxidative Stress in Tissue Sites for Diabetic Complications. Antioxid. Redox Signaling 7, 1543-1552

74. Dills, W. L., Jr. (1993) Protein fructosylation: fructose and the Maillard reaction. Am. J. Clin. Nutr. 58, 779S-787S

75. Suarez, G., Rajaram, R., Oronsky, A. L., and Gawinowicz, M. A. (1989) Nonenzymatic glycation of bovine serum albumin by fructose (fructation). Comparison with the Maillard reaction initiated by glucose. J. Biol. Chem. 264, 3674-3679

76. Schalkwijk, C. G., Stehouwer, C. D., and van Hinsbergh, V. W. (2004) Fructose-mediated non-enzymatic glycation: sweet coupling or bad modification. Diabetes / Metab. Res. Rev. 20, 369-382

77. Bose, T., and Chakraborti, A. S. (2008) Fructose-induced structural and functional modifications of hemoglobin: implication for oxidative stress in diabetes mellitus. Biochim. Biophys. Acta 1780, 800-808

78. Schmidt, A. M., Hori, O., Brett, J., Yan, S. D., Wautier, J. L., and Stern, D. (1994) Cellular receptors for advanced glycation end products. Implications for induction of oxidant stress and cellular dysfunction in the pathogenesis of vascular lesions. Arterioscler. Thromb. 14, 1521-1528

79. Warburg, O. (1956) On the origin of cancer cells. Science 123, 309-314

80. Chan, E. C., Koh, P. K., Mal, M., Cheah, P. Y., Eu, K. W., Backshall, A., Cavill, R., Nicholson, J. K., and Keun, H. C. (2009) Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). J. Proteome Res. 8, 352-361

81. Hirayama, A., Kami, K., Sugimoto, M., Sugawara, M., Toki, N., Onozuka, H., Kinoshita, T., Saito, N., Ochiai, A., Tomita, M., Esumi, H., and Soga, T. (2009) Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. Cancer Res. 69, 4918-4925

82. Vizcaino, J. A., Cote, R., Reisinger, F., Barsnes, H., Foster, J. M., Rameseder, J., Hermjakob, H., and Martens, L. (2010) The Proteomics Identifications database: 2010 update. Nucleic Acids Res. 38, D736-742

5.1.6 SUPPLEMENTARY TABLES

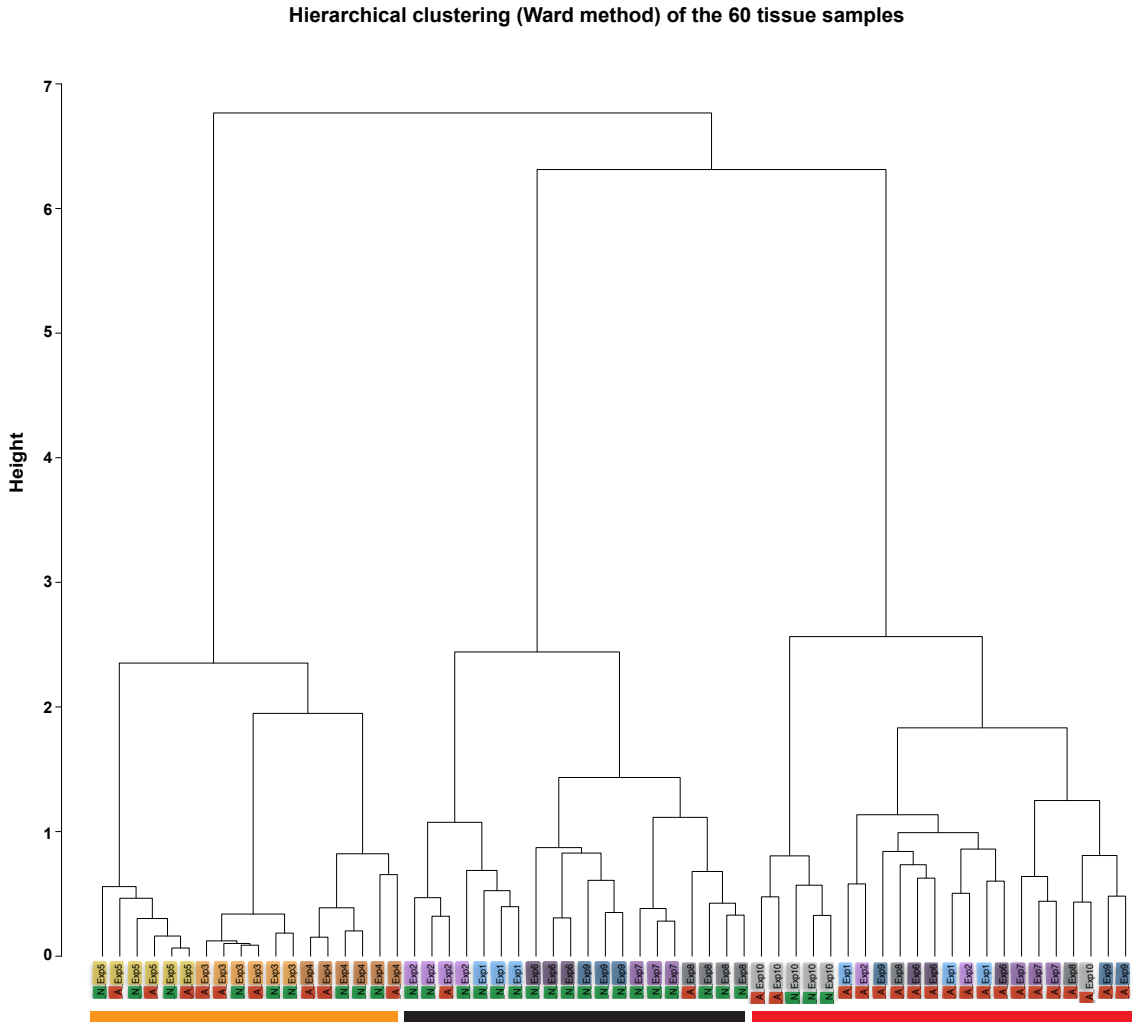
Supplementary table 1 and 3: Due to file size, the links to these tables are available in "Appendix"

Supplementary table 2. Mascot protein search summary for iTRAQ labeling efficiency

iTRAQ experiments on colorectal tissues and in cell lines	PSMs*			iTRAQ labeling efficiency (%)
	Total assigned	With iTRAQ modification (N-term,Y, K)	Without iTRAQ modification (N-term,Y, K)	
Tissue experiment 1	33935	32222	1713	95
Tissue experiment 2	33916	31737	2179	94
Tissue experiment 3	38708	38193	515	99
Tissue experiment 4	41273	40763	510	99
Tissue experiment 5	48340	47532	808	98
Tissue experiment 6	35650	33789	1861	95
Tissue experiment 7	29283	28558	725	98
Tissue experiment 8	28157	27133	1024	96
Tissue experiment 9	32311	30612	1699	95
Tissue experiment 10	33631	30018	3613	89
Cell line Experiment	34961	33361	1600	95

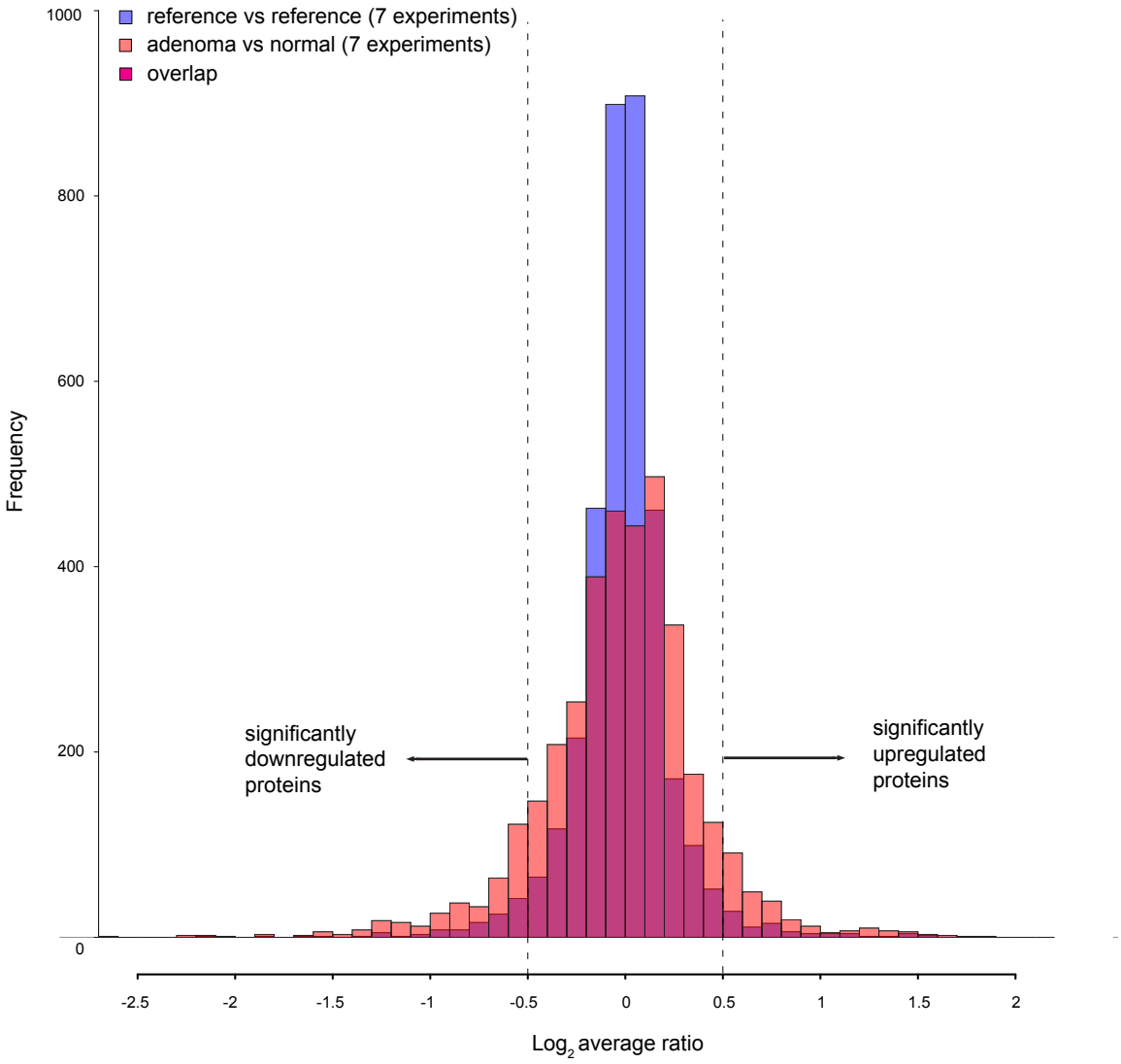
* Assigned Peptide Spectra Matches (PSMs) after exclusion of decoy and contaminant hits.

5.1.7 SUPPLEMENTARY FIGURES



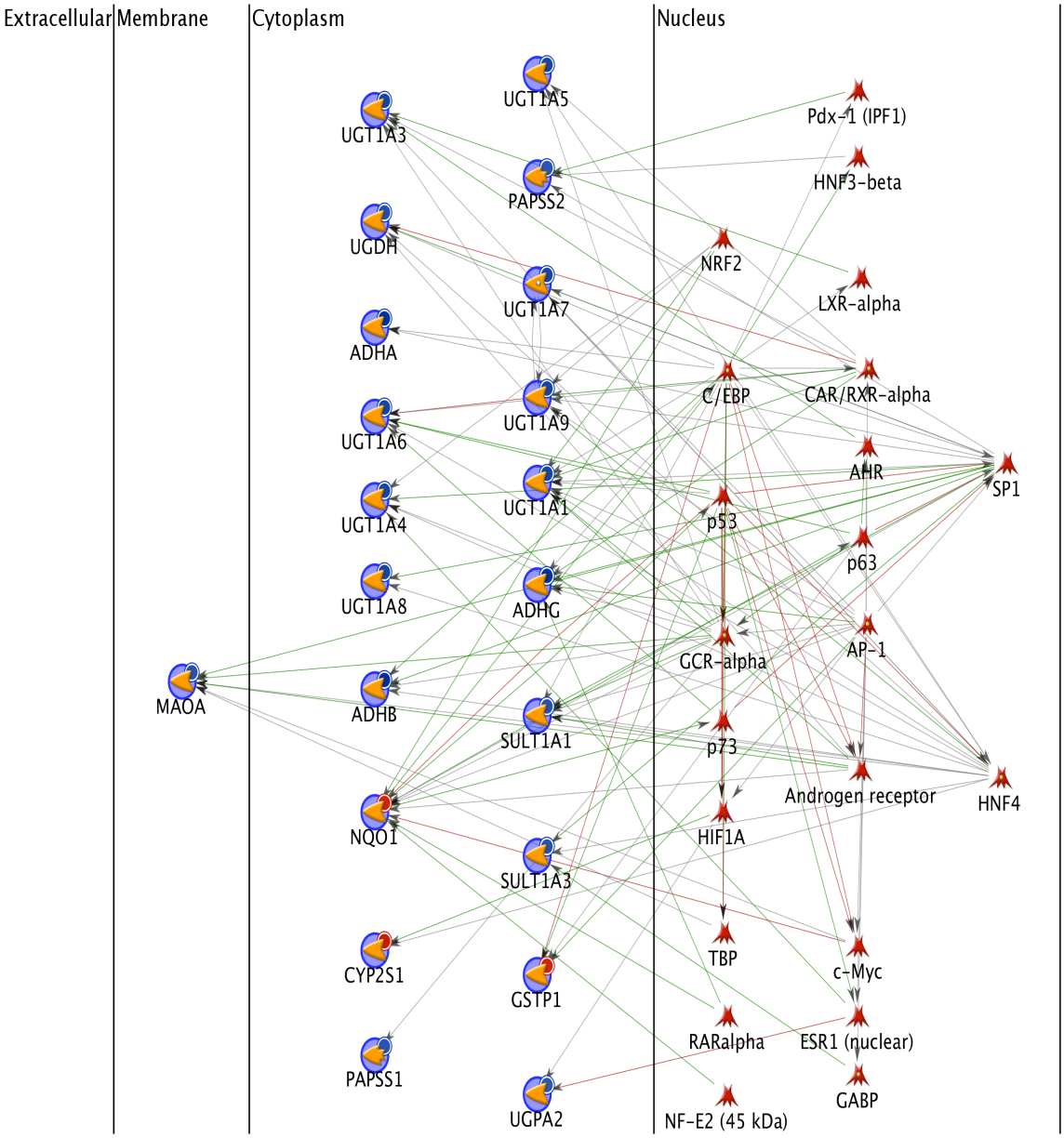
Supplementary figure 1. Hierarchical clustering analysis of colorectal tissue proteomics data (Ward method).

Two distinct clusters emerged: one containing mainly adenomas (A, red bar), the other consisting mostly of normal mucosa samples (N, black bar). The 18 samples in the third cluster (orange bar) comprised nine normal/adenoma tissue pairs, which had been processed in the same labeling and shot-gun run batch. These samples were considered sub-standard since their clustering and sub-clustering appeared to be experiment specific. The samples analyzed in the same experiment (Exp) have the same color



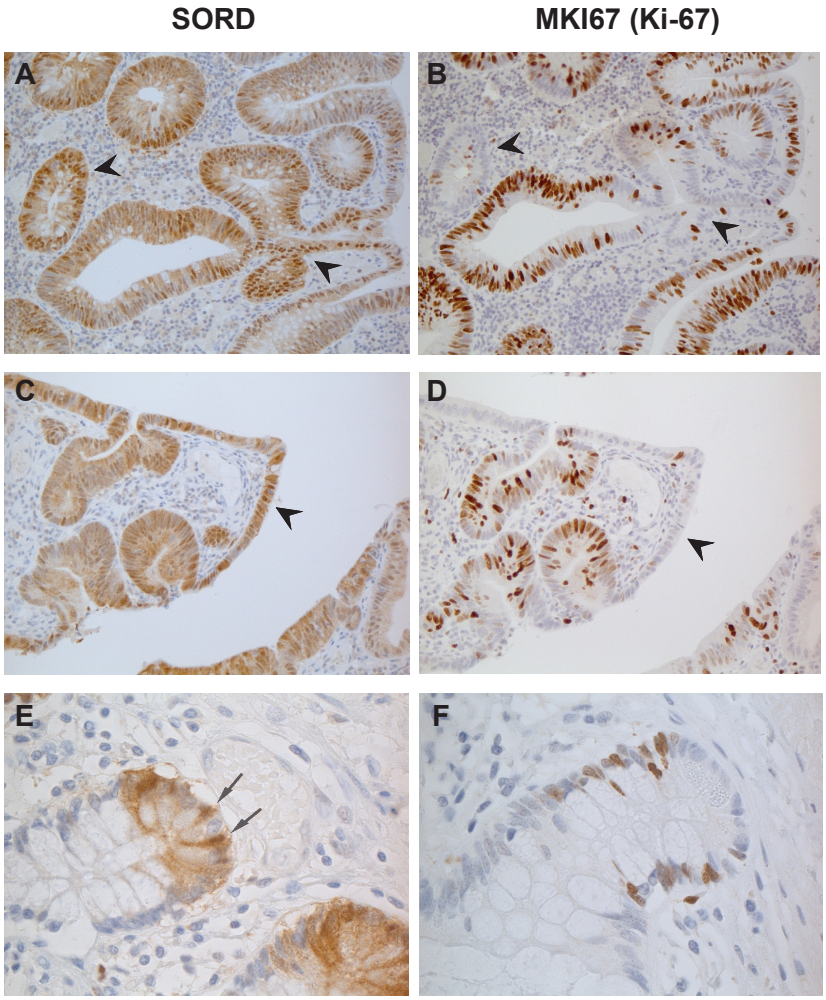
Supplementary figure 2. Significance of protein expression changes.

Histogram comparing average log2 fold changes (adenoma vs. normal, pink bars) for proteins identified in the 21 paired tissue samples (7 iTRAQ experiments) with average $\log_2$ ratios (reference 113:reference 114) for the same proteins in the 7 reference standards. Seventeen percent of the fold changes observed in the tissue samples (vs. only 5.3% of those in the reference standards) exceeded the experimentally defined threshold of ± 0.5 $\log_2$ (± 1.4 in linear scale) (sections indicated with horizontal black arrows).



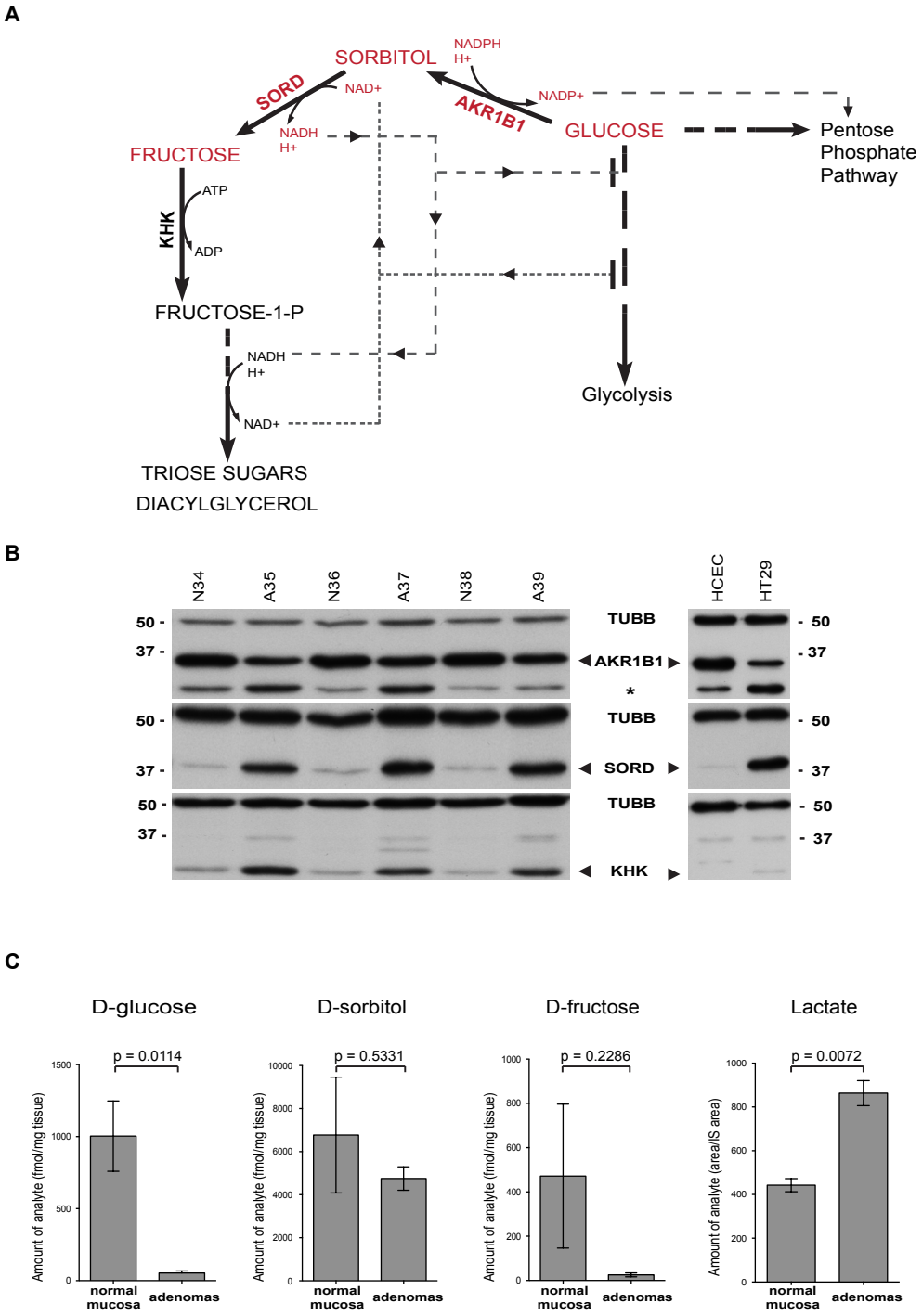
Supplementary figure 3. Network analysis of dysregulated proteins involved in xenobiotic metabolism.

Transcription regulation pathways connecting the 21 proteins listed in the GO xenobiotic metabolism category were investigated using Metacore's Shortest Paths algorithm (GeneGo, Inc., St. Joseph, MI, USA). The transcription factors are represented by red symbols and localized in the nucleus. The functions of the 21 proteins are regulated by SP1 and other known cancer-related transcription factors, such as HIF1A, TP53, and MYC. Although HNF4 had a large number of connections, most of them were unspecified. Biological effects are indicated by green (activation), red (inhibition), and gray (unspecified) lines, with arrows specifying the direction of the interaction. Red and blue circles represent proteins that displayed adenoma-related up- and downregulation, respectively, in our study.



Supplementary figure 4. Nuclear SORD and Ki-67 immunostaining in colorectal adenomas and normal ileal mucosa.

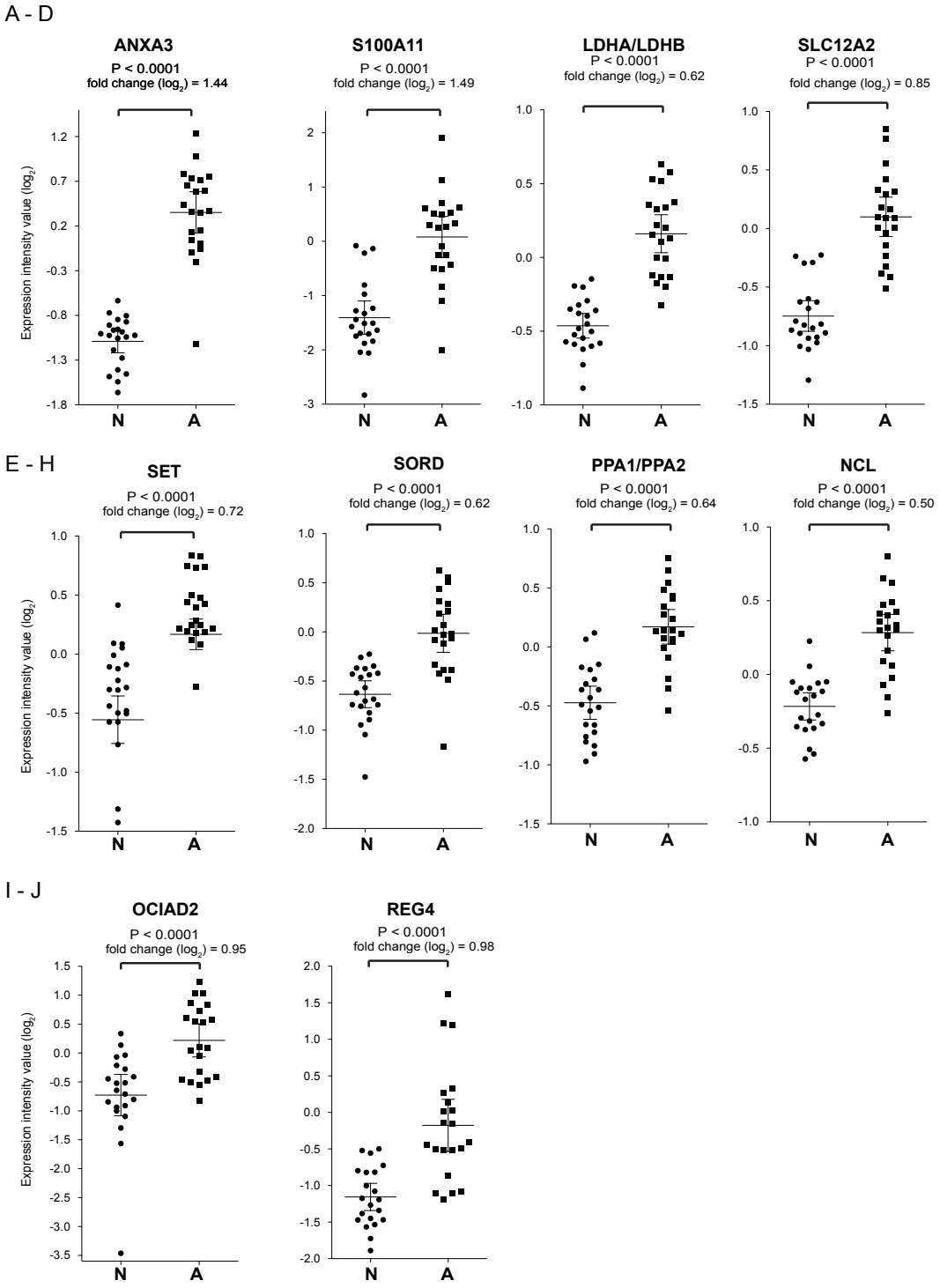
(A and C) Nuclear SORD staining was observed in certain crypts from two different colorectal adenomas, and (B and D) the stained cells were almost invariably negative for Ki-67 nuclear staining (arrowheads). (E) This pattern was also observed in sections of normal ileal crypts, where nuclear SORD staining was limited to the stem cell compartment at crypt bases (arrows; intermingled with Paneth cells), whereas (F) Ki-67 staining was restricted to rapidly proliferating cells in the next compartment, closer to the mouth of the crypt.



Supplementary figure 5. Enzyme expression and metabolite levels in the polyol and related pathways during colorectal tumorigenesis.

(A) The polyol pathway for sorbitol metabolism (depicted in red) and related metabolic pathways (in black). (B) Western blotting documented dysregulated expression of the polyol pathway enzymes, AKR1B1 and SORD, and of KHK in colorectal adenomas, and in HT29 and HCEC cell lines. Specific bands are indicated with arrowheads. The AKR1B1 antibodies (upper panel) detected a second, faster-migrating protein (asterisk) whose levels were also altered in tumor cells but in the opposite direction relative to AKR1B1. The identity of this polypeptide is unknown. (C) Adenomas also presented decreased levels of D-glucose, D-fructose, and D-sorbitol and increased lactate concentrations (vs.

normal mucosa). P values were obtained using two-tailed, paired t-test for total D-glucose (α and β), total D-fructose (α and β), D-sorbitol, and lactate. Columns and error bars indicate means and SEMs for levels measured in three adenoma/normal mucosa pairs.



Supplementary figure 6. Proteins upregulated in adenomas.

Log_2 expression intensity for the 10 proteins displaying the most significant upregulation in adenomas (Table 3). iTRAQ measured protein abundance ratios in the normal mucosa (N) and corresponding adenomatous lesions (A) of patients is indicated with black dots and squares respectively. The error bars represent the mean protein abundance and 95% confidence intervals.

5.2 An optimized, large-scale SRM method for the detection and verification of early markers of colorectal tumorigenesis

The second aim of my project was to develop and optimize a large-scale selected reaction monitoring (SRM) workflow to verify candidate colorectal tumor markers (identified from aim 1).

5.2.1 INTRODUCTION

Non-hereditary colorectal cancer is ranked as one of the major causes of high cancer-related morbidity and mortality rates in adult individuals (Siegel et al., 2014; 2013). Most of these sporadic forms of colorectal cancer occur along an adenoma-adenocarcinoma-metastasis route (Fearon and Vogelstein, 1990). They begin with the formation of benign but precanceorus lesions of the colorectal mucosa (i.e. adenomas). With time, these neoplasms progresses from large adenomas, through malignant stage I-IV cancers to metastatic conditions (Anon, 2003; Cattaneo et al., 2010; Fearon and Vogelstein, 1990). Adenomatuous lesions carry genetic and epigenetic alterations, especially in the Wnt-APC-βcatenin pathway, which dysregulate normal cell proliferation-differentiation-apoptosis homeostasis, yet do not invade the underlying submucosa. They could however acquire further mutations with time, and advance to the cancerous state. Disease prognosis and treatment are highly positive when tumors are detected at the adenoma or early adenocarcinoma stage (Cunningham et al., 2010; de Wit et al., 2012; Siegel et al., 2014), facilitating their complete removal. Therefore, the early detection of neoplasms remains the best way to curb the incidence of colorectal cancer. Limitations in available screening techniques have hampered major progress in curbing this incidence (Ang et al., 2011; Jimenez et al., 2010; Luo et al., 2013). Endoscopic techniques are invasive, costly and time-consuming, although they are highly precise diagnostically and allow prompt removal of early lesions. On the other hand, faecal-based tests have low specificity and have been reported to be more effective in detecting advanced tumors (Ang et al., 2011; Davies et al., 2005; Song et al., 2004).

Targeted proteomic methods have improved in leaps and bounds over the last years (Liu et al., 2013a; 2013b). Some of the reasons for this include improved sensitivity, accuracy and reproducibility of these methods over data dependent techniques. Selected reaction monitoring (SRM), also known as multiple reaction monitoring (MRM), is a highly reproducible and sensitive targeted proteomic technique for the quantification of specific proteins in a complex sample background (Lange et al., 2008; Picotti et al., 2009). An SRM workflow involves the monitoring of predetermined peptide coordinates over time to detect and precisely quantify protein abundance. These peptide coordinates include selected precursor ions and their corresponding fragment ions- collectively called transitions, as well as the elution time (retention time) of each transition. SRM experiments are performed in triple quadrupole instruments. While the first and third quadrupoles are set as filters to precisely choose pre-

determined m/z values for the peptide precursor and corresponding fragment ions respectively, the second quadrupole acts as a collision cell where precursor ions are fragmented at an optimal collision energy. In a single SRM experiment, numerous precursor/fragment ion pairs (i.e., transitions) can be monitored across a relatively large number of samples. The resulting chromatographic trace comprises information on the retention time and signal intensity per transition, and is used to determine target peptide abundance. With the addition of known concentrations of high-purity heavy isotope-labelled reference peptides to the test sample, absolute peptide abundance can be measured. More so, protein abundance can be compared in various samples and test conditions since peptide abundance is measured relative to a defined concentration of the reference standard (Lange et al., 2008; Parker and Borchers, 2014; Picotti and Aebersold, 2012). Advances in SRM technology have recently incorporated new statistical tools to support the design of SRM experiments and the analysis of large quantitative data sets generated from these experiments (Chang et al., 2012; Choi et al., 2014; Hüttenhain et al., 2009; 2012; MacLean et al., 2010; Reiter et al., 2011; Selevsek et al., 2011; Surinova et al., 2013). For example, open-source softwares such as *Skyline* and *MSstats* are of benefit in both upstream SRM assay development and refinement, and downstream analysis to determine significant protein changes across multiple conditions (Chang et al., 2012; Choi et al., 2014; MacLean et al., 2010). Also, the implementation of automated, high-confident peptide-scoring algorithms such as *mProphet* has greatly improved the accuracy of protein abundance measurements by SRM (Reiter et al., 2011).

Although numerous candidate cancer biomarkers have been proposed based on proteomic studies on colorectal cancer, the on-going search for diagnostic clinical biomarkers for this disease is yet to yield promising results (de Wit et al., 2012; Füzéry et al., 2013; Jimenez et al., 2010; Luo et al., 2013). This is mainly due to the arduous path governing the translation of biomarker discovery to clinical application and the lack of appropriate tools to identify and reproducibly quantify biomarker candidates over multiple samples.

The pipeline from biomarker discovery to clinical use involves clearly defined phases, discussed extensively in (Rifai et al., 2006), (Kulasingam et al., 2010), and (Surinova et al., 2011). In a discovery-based study, researchers comprehensively study the entire proteome, genome or metabolome and aim to discover interesting alterations characteristic to the diseased state (Domon and Aebersold, 2006; 2010). Tissue samples are usually employed in most discovery phase studies because the

concentration of candidate biomarkers is usually highest at the disease site. More over, tumor-associated proteins in tissues could subsequently be shed or secreted into the blood stream. The preclinical verification phase serves as a bridge between biomarker discovery and clinical validation. Biomarker verification entails the development of sensitive, reproducible assays to precisely measure protein abundance changes in a large number of well-defined sample sets that are representative of the target population. The rationale here is to specifically target the tissue-derived proteins of interest in large cohorts of plasma samples to determine its clinical and analytical performance (Meng and Veenstra, 2011; Surinova et al., 2011). In the validation phase, standardized assays are used to determine the performance of the biomarker panel and to define the criteria for its application in the clinical evaluation of the disease (Surinova et al., 2011). Patient groups are designed retrospectively and the sample cohorts are much larger than in the verification phase. The next phase, clinical evaluation aims to employ approved state of the art technology to translate the outcome of the preclinical investigations to the clinics. And finally in the disease control phase, the impact of the biomarker test on the disease burden in the target population is assessed.

Immuno-based tests such as enzyme-linked immunosorbent assays (ELISA) remain the preferred preclinical verification technique. In comparison to SRM, ELISA is an antibody-dependent technique, which is not ideal when a large number of tumor biomarkers are to be tested (Issaq and Veenstra, 2008; Parker and Borchers, 2014). Although recent reports show positive correlations between SRM and ELISA protein concentration levels, SRM technique was more effective in detecting protein variants and isoforms (Parker and Borchers, 2014; Shi et al., 2014). There are huge expectations in the proteomics community that targeted proteomic techniques would be adopted as standard complementary methods for the clinical verification and validation of candidate disease biomarkers. Besides, an increasing number of proteomic-based cancer studies have incorporated SRM for the verification of candidate biomarkers (Hüttenhain et al., 2009; Lopez et al., 2011; Muraoka et al., 2012; Parker and Borchers, 2014; Selevsek et al., 2011). These biomarkers are most commonly identified from a preliminary discovery phase study.

Candidate protein markers for colorectal cancer have been sourced largely from shotgun data (discovery phase), or predicted based on information from other -omic studies (genetic, epigenetic and transcriptomic studies). However, most proteomic data sourced this way have been largely from the *cancerous* lesions. We have earlier

reported a comprehensive quantitative shotgun proteomic study on a relatively large number of human colorectal precancerous lesions and their matching normal mucosa samples (Uzozie et al., 2014). In addition to the novel protein alterations detected in these tissues, we observed that over 90% of proteins with adenoma-related upregulation have been reported in literature to be associated with tumorigenesis. In continuation of our aim to verify putative diagnostic markers representing tumor stages along the adenoma-adenocarcinoma route, we proceeded to establish a reproducible, targeted assay to profile candidate protein markers from this study in an independent large group of colorectal neoplasms, comparative to the normal mucosa. We generated reliable SRM assays to uniquely and confidently quantify target proteins differentially expressed in early (adenomas) and advanced (adenocarcinomas) colorectal neoplasms. In addition to verifying putative tumor biomarkers for colorectal tumors, we describe a comprehensive and reproducible SRM workflow that would systematically aid the quantitative analysis of potential tumor markers across large cohorts of patient plasma samples.

5.2.2 EXPERIMENTAL PROCEDURES

5.2.2.1 Human tissue samples and cell lines

The local ethics committee approved the use of tissue samples, and all samples were used in agreement with the Declaration of Helsinki. Human colorectal tissues were prospectively collected from patients during colonoscopy, at the Instituti Ospitalieri of Cremona, Italy. A coding system was applied to each donor to protect human confidentiality. All donors provided written consent to sample collection, testing, and data publication. The sample series comprised 19 adenomas and 17 adenocarcinomas along with their corresponding normal mucosa (total, 72 tissue samples (Table 5)**.** Each paired normal sample was collected from the same colon segment where the tumor was located, at >2 cm from the lesion. Samples were immediately frozen in liquid nitrogen and stored at -80 ^{o}C.

Table 5. Characteristics of the patients with neoplastic lesions included in the study.

Patient number	Age	Sex	Colon segment involved	Maximum lesion diameter (mm)	Paris classification #	Pit pattern classification °	Microscopic appearance	Highest degree of dysplasia in the lesion ▪	No. of lesions at study colonoscopy ∞	No. of previously excised lesions ‡
1	50	F	C	30	0 - IIa	IIIL	TA	LGD	1	1
2	72	F	A	20	Is	IV	TVA	LGD	1	0
3	66	F	S	35	Ip	IV	TA	LGD	1	0
4	78	M	R	50	IIa + IIs	IV + Vi	TVA	LGD	1	0
5	58	F	R	20	IIa+IIc	IIIL -Vi	TA	LGD	2	0
6	60	M	C	30	Is	IV	TVA	LGD	2	2
7	65	M	A	40	IIa + Is	IV	TA	LGD	1	0
8	70	F	R	20	Is + IIa	IV	TVA	LGD	1	0
9	70	M	A	40	Is + IIa	IV	TA	LGD	1	0
10	77	F	R	80	IIa+Is	IV	VA	LGD	1	0
11	83	M	S	20	IIa+Is	IIIL	TVA	LGD	3	0
12	75	F	S	30	IIa+Is	IIIL -IV	TA	LGD	1	3
13	63	M	T	30	0 - Is	IV	TVA	LGD	1	0
14	74	F	A	80	Is + IIa	IV	TVA	LGD	1	0
15	82	M	D	50	Is + IIa	IV	AV	LGD	1	0
16	61	F	R	40	Is +IIa	II	SSA	LGD	1	0
17	78	M	R	60	IIa + Is	Vi - IV	TA	HGD-ADK	1	1
18	72	F	C	45	0 - Is	IV	TA	HGD	1	0
19	58	M	D	20	0 - Ip	IV	TA	HGD	2	0
20	83	F	C	35	Is	IV	TVA	HGD	1	0
21	58	M	S	40	Is	IV + Vi	TVA	ADK	2	1
22	67	F	R	40	0 - Is	Vn	(ADK)	ADK	1	0
23	37	M	R	60	0 - Is	Vn	(ADK)	ADK	1	0
24	63	M	S	80	0 - Is	Vn	(ADK)	ADK	1	0
25	82	M	S	65	0 - Is	Vn	(ADK)	ADK	1	0
26	52	M	S	60	0 - Is	Vn	(ADK)	ADK	1	0
27	56	F	A	65	0 - Is	Vn	(ADK)	ADK	1	0
28	55	M	R	70	0 - Is	Vn	(ADK)	ADK	3	0
29	87	F	T	70	0 - Is	Vn	(ADK)	ADK	1	0
30	64	F	A	60	0 - Is	Vn	(ADK)	ADK	1	0
31	64	M	C	70	Is	NA	(ADK)	ADK	1	0
32	88	F	A	55	0 - Is	Vi	(ADK)	ADK	1	0
33	67	M	T	80	0 - Is	Vn	(ADK)	ADK	1	0
34	64	M	R-S	100	0 - Is	Vn	(ADK)	ADK	2	1
35	61	F	T	60	0-Is	Vn	(ADK)	ADK	1	0
36	53	F	A	50	0-Is	Vn	(ADK)	ADK	1	0

Abbreviations: M, male; F, female; C, caecum; A, ascending colon; T, transversum; D, descending colon; S, sigma; R, rectum; TA, tubular adenoma; TVA, tubulovillous adenoma; VA, villous adenoma; SSA, sessile serrated adenoma; LGD, low-grade dysplasia; HGD, high-grade dysplasia; ADK, Adenocarcinoma; NA, data not available

Macroscopic appearance of neoplastic lesions was classified according to Paris Endoscopic Classification. The Paris Endoscopic Classification of Superficial Neoplastic Lesions. Gastrointest Endosc 2003;58(suppl.):S3-S27

° Morphological analysis of colonic crypt patterns according to Kudo's classification. Kudo S, Rubio CA, Teixeira CR, et al. Pit pattern in colorectal neoplasia:endoscopic magnifying view. Endoscopy 2001;33;367-7

▪ Low-grade versus high-grade dysplasia as defined by the WHO classification of tumors of the digestive system, editorial and consensus conference in Lyon, France, November 6-9, 1999.IARC

∞ This number includes the lesion subjected to the proteomic study

‡ Total no. adenomas detected and excised during previous colonoscopies.

5.2.2.2 Protein extraction from tissues

Frozen tissue samples were quickly weighed and homogenized on ice in a Wheaton glass borosilicate grinder containing a solution of 100 mM triethylammonium bicarbonate (Sigma, St Louis, MO, USA), 1X Complete EDTA-free Protease Inhibitor Cocktail (Roche, Mannheim, Germany), 1 M Urea, 5 mM β-glycerophosphate disodium salt hydrate, 1 mM sodium orthovanadate, and 5 mM sodium fluoride (Sigma, St Louis,

MO, USA). Homogenates were then sonicated with a Bioruptor (Diagenode, Denville, NJ, USA) (high power, five 10"/10" ON/OFF cycles) and centrifuged (16,000 *g* for 5 mins at 4 ^{o}C). The resulting supernatant containing proteins was collected and stored at -80 ^{o}C.

5.2.2.3 Protein digestion

For each sample, 50 micrograms of protein was used. Proteins were reduced in denaturing buffer with 10 mM DTT at 35^{o}C for 45 min, and alkylated with 50 mM iodoacetamide for 1 h at room temperature and in the dark. To terminate the alkylation reaction, 50 mM DTT was added, and samples were incubated for 10 min at room temperature. Samples were further incubated in sequencing grade endoproteinase Lys-C (Roche, Mannheim, Germany) at a final enzyme to substrate ratio of 1:100 for 3 hrs at 23^{o}C, with constant shaking, and then diluted with 50 mM ammonium bicarbonate to a final concentration of 1.4 M urea. Trypsin (sequencing-grade modified trypsin; Promega, Madison, WI, USA) digestion was performed at a final enzyme to substrate ratio of 1:50, for 15 hrs. Digestion was stopped by adding a solution of 30% acetonitrile, 1% trifluoroacetic acetic to a final concentration of 3% acetonitrile, and 1% trifluoroacetic acid. Peptide solutions were desalted on 100 mg SepPak C18 column (Finisterre, Teknokroma, Barcelona, Spain). Briefly, each column was wet once with 1000 µL of 100% methanol, washed once with 1000 µL of 80% acetonitrile, and equilibrated twice with a solution of 3% acetonitrile and 0.1% trifluoroacetic acid. Acidified digests were then loaded and the column was washed twice with 1000 µL of a solution of 3% acetonitrile and 1% trifluoroacetic acid. Elution was done once with 500 µl of 60% acetonitrile and 0.1% trifluoroacetic acid. Peptides were lyophilized in a vacuum centrifuge and resolubilized in 50 µl (1:1 v/w) of a solution of 3% acetonitrile, 0.1% formic acid, and iRT 10x mix (iRT mix was prepared according to the manufacturer's protocol; Biognosys AG, Zurich, Switzerland)

5.2.2.4 Selection of target proteins and endogenous (light) proteotypic peptides

A schematic of the SRM workflow is detailed in Figure 14. The proteins of interest were selected based on the results of a quantitative (iTRAQ 8plex) shotgun study on human colorectal adenomatous lesions, reported in (Uzozie et al., 2014). The selection comprised proteins of the following categories: [1] those with adenoma-related dysregulation reported in (Uzozie et al., 2014), [2] those with adenoma-related dysregulation reported in (Cattaneo et al., 2011), [3] global standard control proteins (also known as housekeeping proteins), [4] proteins predicted to be involved in

pathways that [1] are involved in. A total of forty candidates, listed in Table 6**,** were chosen for verification by SRM.

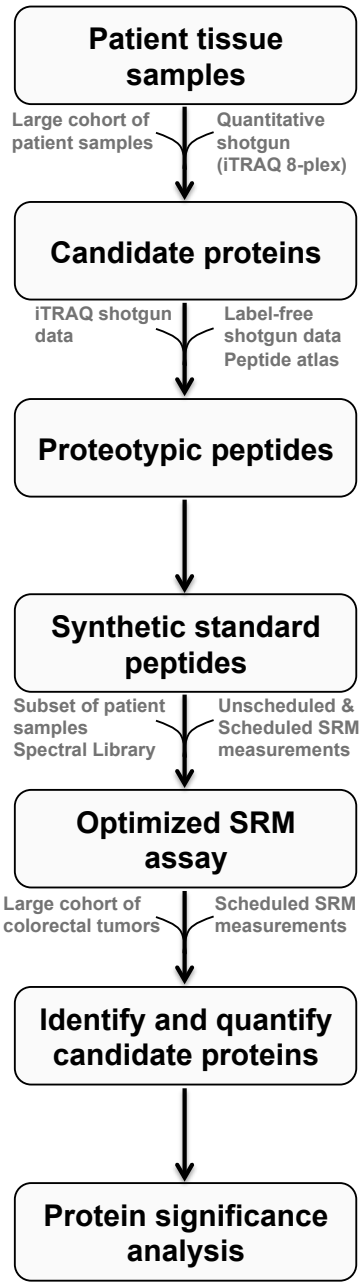


Figure 14. Selected reaction monitoring (SRM) workflow employed in this study.

The different steps from candidate protein selection through SRM assay development and refinement, and analysis of scheduled SRM data to identify proteins with significant protein changes in the tissue samples studied (see Experimental Procedures for details).

For each target protein, the corresponding proteotypic light peptides for SRM were sourced in three ways (detailed in Figure 14) - [1] iTRAQ 8plex-based quantitative shotgun data from our previously reported study (Uzozie et al., 2014), [2] label-free shotgun data on a subset of paired tissue samples, and [3] data in Peptide Atlas. Each

proteotypic peptide selected was unique for the specified target protein. One hundred forty peptides were selected for the 40 protein targets (Table 6). The label free shotgun study was necessary to confirm peptide detection in tissue samples because no fractionation technique was included for the SRM study in comparison to our previous iTRAQ study (Uzozie et al., 2014). Briefly, label-free shotgun analysis was performed in an LTQ Orbitrap Velos, (Thermo Fisher Scientific, Bremen, Germany) on 1 microgram of peptide sample. The acquired spectra were searched against a human protein database. Peptides with a protein FDR of <1% were selected for further use.

Table 6. List of 40 proteins and corresponding 140 proteotypic peptides for SRM assay development.

Uniprot Accession No.	Protein Name	Peptide Sequence
O60218 [4]	AK1BA	HIDCAYVYQNEHEVGEAIQEK
		LSDEEMATILSFNR
		ALGVSNFSHFQIEK
O60575 [1]	ISK4	QWVIALALAALLVVDR
		EVPVAAGK
		QDIQIMK
O75531 [1]	BAF	DFVAEPMGEKPVGSLAGIGEVLGK
		AYVVLGQFLVLK
		DTCGANAK
O76041 [1]	NEBL	TDPGSIFDLDPLEDNIQSR
		TTQQNISAVFYK
		GVHPHIVEMDR
		NIGAFISEAK
P00338 [1]	LDHA	DQLIYNLLK
		DLADELALVDVIEDK
		QVVESAYEVIK
P04406 [3]	G3P	GALQNIIPASTGAAK
		VPTANVSVVDLTCR
		LVINGNPITIFQER
		LISWYDNEFGYSNR
P07195 [1]	LDHB	LIAPVAEEEATVPNNK
		SLADELALVDVLEDK

		IVVVTAGVR
P07900 [3]	HS90A	NPDDITNEEYGEFYK
		HLEINPDHSIIETLR
		DQVANSAFVER
		APFDLFENR
P11413 [4]	G6PD	NSYVAGQYDDAASYQR
		LQFHDVAGDIFHQQCK
		GGYFDEFGIIR
		VGFQYEGTYK
P12429 [1]	ANXA3	DYPDFSPSVDAEAIQK
		SDTSGDYEITLLK
		HYGYSLYSAIK
		LTFDEYR
P15121 [4]	ALDR	SPPGQVTEAVK
		MPILGLGTWK
		VAIDVGYR
		TTAQVLIR
P19338 [1]	NUCL	FGYVDFESAEDLEK
		EVFEDAAEIR
		GFGFVDFNSEEDAK
P21695 [4]	GPDA	IVGGNAAQLAQFDPR
		LISEVIGER
		ELYSILQHK
P22223 [2]	CADH3	EAEVTLEAGGAEQEPGQALGK
		DWVVAPISVPENGK
		STGTISVISSGLDR
		YEAHVPENAVGHEVQR
		LTVTDLDAPNSPAWR
		GLEARPEVVLR
P25815 [1]	S100P	YSGSEGSTQTLTK
		ELPGFLQSGK
P27695 [1]	APEX1	EAAGEGPALYEDPPDQK
		QGFGELLQAVPLADSFR
		LDYFLLSHSLLPALCDSK

P31949 [1]	S10AB	ISSPTETER
		DGYNYTLSK
		DPGVLDR
P35557 [4]	HXK4	LVDESSANPGQQLYEK
		ASGAEGNNVVGLLR
		LPLGFTFSFPVR
P36952 [1]	SPB5	DLTDGHFENILADNSVNDQTK
		GDTANEIGQVLHFENVK
		ILVVNAAYFVGK
		IIELPFQNK
P39687 [1]	AN32A	ELVLDNSR
		CPNLTHLNLSGNK
P50053 [4]	KHK	EELFQLFGYGDVVFVSK
		HLGFQSAEEALR
		LLHSDAFPPPR
P55011 [1]	S12A2	VELPGTAVPSVPEDAAPASR
		AFYAPVHADDLR
		EGLDISHLQGQEELLSSQEK
		ITDNELELYK
P62330 [1]	ARF6	QDLPDAMKPHEIQEK
		ILMLGLDAAGK
		FNVWDVGGQDK
P80188 [1]	NGAL	TFVPGCQPGEFTLGNIK
		VPLQQNFQDNQFQGK
		WYVVGLAGNAILR
		SYPGLTSYLVR
Q00796 [1]	DHSO (SORD)	LENYPIPEPGPNEVLLR
		KPMVLGHEASGTVEK
		VAIEPGAPR
		YNLSPSIFFCATPPDDGNLCR
		LPDNVTFEEGALIEPLSVGIHACR
		AMGAAQVVVTDLSATR
		EIGADLVLQISK
Q01105 [1]	SET	VEVTEFEDIK

		IDFYFDENPYFENK
		EFHLNESGDPSSK
Q15124 [1]	PGM5	NYLPNFIQSVLSSIDLR
		DGLWAVLVWLSIIAAR
		VPVYETPAGWR
Q15181 [1]	IPYR	AAPFSLEYR
		GQYISPFHDIPIYADK
		DVFHMVVEVPR
		DPLNPIK
Q15293 [1]	RCN1	HWILPQDYDHAQAEAR
		TFDQLTPDESK
		YIFDNVAK
Q15582 [1]	BGH3	EGVYTVFAPTNEAFR
		GDELADSALEIFK
		LTLLAPLNSVFK
Q56VL3 [1]	OCAD2	VALAGLLGFGLGK
		GAGFGPQHNR
		FHFFEDQLR
		DAHFPPPSK
Q6UX06 [1]	OLFM4	VNLTTNTIAVTQTLPNAAYNNR
		GLYWVAPLNTDGR
		GFSYLYGAWGR
Q8WUJ3 [2]	K1199	TLHPGGMAEGGYFFER
		HPWSFLTVK
		IWGPGGLDHSGR
		TLPIGQNFPIR
		STHYQQYQPVVTLQK
		IFQVVPIPVVK
Q92688 [1]	AN32B	ELVLDNCK
		LPNLTHLNLSGNK
Q96CX2 [1]	KCD12	LGAPQQPGPGPPPSR
		DLQLVLPDYFPER
		EAEYFELPELVR
		FNFLEQAFDK

Q96IF1 [2]	AJUBA	GATGGPGDEPLEPAR
		DYHFECYHCEDCR
		YQDELTALLR
		SYHPGCFR
Q9BYZ8 [1]	REG4	SNCYGYFR
		EASTIAEYISGYQR
Q9C002 [1]	NMES1	LITINQQWKPIEELQNVQR
		NPEPWETVDPTVPQK
		TDVILDR
Q9H2U2 [1]	IPYR2	NVTGHYISPFHDIPLK
		ILGILALIDEGETDWK
		LIAINANDPEASK
		AFALEVIK
Q9UKY7 [1]	CDV3	LQLDNQYAVLENQK
		VQAMQISSEK
		SLDNFFAK

[1] Proteins with adenoma-related dysregulation reported in (Uzozie et al., 2014).

[2] Proteins with adenoma-related dysregulation reported in (Cattaneo et al., 2011).

[3] Global standard control proteins (also known as housekeeping proteins).

[4] Proteins predicted to be involved in pathways that [1] are involved in.

5.2.2.5 Synthetic isotope-labelled (heavy) reference peptides for SRM assay library

Heavy isotope-labelled reference peptides were purchased from JPT Peptide Technologies GmbH (Berlin, Germany). Each heavy peptide had an identical sequence to the corresponding light peptide, but its C-terminal lysine or arginine was fully labelled with ^{15}N,^{13}C respectively. A pool of all heavy peptides (with known concentration of each peptide) in the sample matrix solution was prepared and analyzed on an LTQ Orbitrap Velos (Thermo Fischer Scientific, Bremen, Germany) by nanoLC-MS/MS. Mascot output .dat files were imported into Skyline and used to generate the spectral library for creation of SRM transitions. These high purity peptides were also used to optimize our SRM assay.

5.2.2.6 SRM assay development and scheduled SRM analysis

SRM assay development was performed on a TSQ Vantage Triple Quadrupole Mass Spectrometer (Thermo Fischer Scientific, Bremen, Germany), using a sample mixture containing known amounts of each heavy isotope-labelled peptide and unscheduled SRM methods generated from the spectral library. A preliminary SRM-transition list was created from the spectral library in Skyline and up to six fragment ions were selected for each precursor ion.

Chromatograms from unscheduled SRM runs were imported into Skyline, and a maximum of four best performing transitions was selected for each peptide. Retention times from unscheduled runs were extracted and used to calculate iRT values (Escher et al., 2012) for each peptide in Skyline.

For scheduled SRM runs, iRT peptides were used to calibrate the retention time of each target peptide (Escher et al., 2012). An initial scheduled SRM method was then exported for parameter optimization. Individual parameters such as light peptide contamination of reference mixture, collision energy, dwell time, and predicted retention time were optimized for each transition. Retention time scheduling for all transitions was done according to Skyline scheduling predictions. A window of ± 2.5 min was used to schedule all transitions. The finalized transition list consisted of 600 transitions. Each protein had *at least* one proteotypic peptide and two to four transitions. To aid the detection of low abundant peptides, without comprising the dwell time of 0.020 s per transition, the 600 transitions were split in two scheduled methods based on proteins. The two complete transition lists, including information on peptide sequence Q1, Q3, CE and RT window are available in Supplementary table 1 (Method 1 transitions) and Supplementary table 2 (Method 2 transitions) respectively. SRM transitions for the two global standard peptides and iRT peptides were monitored in each method, while target proteins were monitored in either the first or the second method.

5.2.2.7 High-resolution liquid chromatography-tandem mass spectrometry (LC-MSMS)

Scheduled SRM measurements were performed on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Fischer Scientific, Bremen, Germany), equipped with nanoelectrospray iononization source, and coupled to a NanoLC Ultra 2D HPLC system (Eksigent Technologies, Dublin, CA, USA). The instrument method was set to SRM acquisition mode, and with the following parameters: a Q1 resolution of 0.7 fwhm, a Q3 resolution of 0.7 fwhm, cycle time of 2 s, and dwell time of 0.020 s. A

spray voltage of 1.3 keV was used with a heated ion transfer tube set at a temperature of 270 ^{o}C. Argon was used as the collision gas at a nominal pressure of 1.5 mTorr. Chromatographic separation of peptides was performed on an in-house-made frit column (150 mm x 75 µm) coupled to a fused silica emitter (100 mm x 75 µm) (New Objective Inc., Woburn, MA, USA). Frit columns were packed with reverse-phase C18 material (AQ, 3 µm 200 Å, Bischoff, GmbH, Leonberg, Germany) and maintained at 50^{o}C with an automatic heater during all SRM experiments. Peptides were loaded on the column from a cooled (4^{o}C) Eksigent autosampler and separated with a linear gradient of acetonitrile/water, containing 0.1% formic acid, at a flow rate of 300 nL/min. An elution gradient from 5% to 35% acetonitrile in 60 min was used. Four µL of sample, corresponding to 1.25 µg peptide mass was injected. To avoid sample carryover, blank samples were measured between SRM measurements of paired biological samples.

5.2.2.8 Sample randomization and blocking

A block-randomized strategy (Oberg and Vitek, 2009) was employed in our experimental design. Based on the experiment phases illustrated in Figure 15**,** adenoma and adenocarcinoma pairs were randomly assigned and a balanced allocation of each group was maintained both at the sample preparation and SRM analysis phases.

Protein extraction from samples was randomized based on the tumor type. For the 72 samples investigated, total protein from the same number of normal/adenoma pair and normal/adenocarcinoma pair were prepared per extraction experiment. The digestion of protein extracts from the 72 samples was performed in three groups. One random sample pair from each group was digested twice. This resulted in an additional 6 samples, and a total of 78 samples for scheduled SRM analysis. The order of spectral acquisition on the TSQ Vantage was also randomized for each SRM method (Figure 15). Sample queue for each method was created using the random function in excel and randomization was done based on protein digestion group, protein extraction pairs, day of sample processing and SRM data acquisition. Each patient sample was analyzed twice for transitions in method 1 and 2, and both SRM methods were performed separately. Statistical analysis of SRM data targeted the differences in adenoma vs. normal mucosa group and adenocarcinoma vs. normal mucosa group.

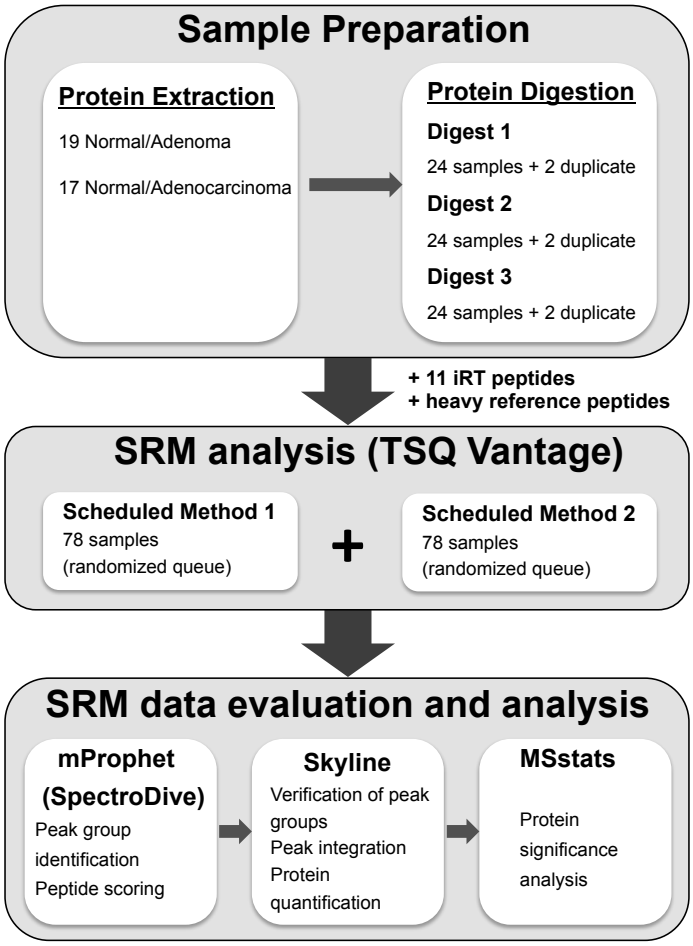


Figure 15. Description of sample preparation and scheduled SRM analysis

One random sample was digested twice in each group (72 to 78 samples total). SRM transitions in Method 1 and Method 2 were monitored in each of the 78 samples spiked with iRT and heavy isotope-labelled peptides according to a randomized SRM sequence. Evaluation and analysis of the SRM spectra was sequencially achieved with mProphet, Skyline and MSstats. (see Experimental procedures for details).

5.2.2.9 Data analysis

SRM data analysis was performed according to the scheme shown in Figure 15. Peak group identification and peptide scoring were actualized using SpectroDive software (Version 5.5.5478.20997, Biognosys AG, Zurich, Switzerland). The data files containing SRM spectra acquired for each sample during the scheduled runs, and the transition assay list containing specified parameters (see Supplementary table 3) served as input files for this analysis. SpectroDive processes the SRM data using the assay list of target and reference peptides, and applies extraction and scoring strategy incorporated in mProphet (Reiter et al., 2011). mProphet is a fully automated system that computes accurate error rates for the identification of target peptides in SRM data

sets and maximises specificity and sensitivity by combining relevant features such as coelution of endogeneous and reference transitions, peak shape and peak intensities, in a probabilistic statistical model. The top-ranked peak groups for each target and decoy transition group are then used to estimate the false discovery rate (FDR) as described by Reiter *et al*., (Reiter et al., 2011) and Cscore, as implemented in SpectroDive. Only peptides identified with a false discovery rate below 1% and a Cscore above 10 in the resulting file from SpectroDive were further processed for quantification (Supplementary table 4).

The output file from SpectroDive (Supplementary table 4), with identified peak groups was reviewed in Skyline to confirm transitions, integrate peak groups and perform protein quantification. Peptides identified in less than 6 samples based on mProphet and Skyline visual analysis were excluded from further processing. Heavy and light peak area intensity for all transitions were exported from Skyline and used to determine proteins with significant abundance differences in adenomas vs. normal mucosa and adenocarcinoma vs. normal mucosa.

Protein significance analysis was performed with MSstats (MSstats.daily 2.1.6) in R statistical package (Chang et al., 2012; Choi et al., 2014). The input file for MSstats analysis contained values for *Conditon*, *Bioreplicate*, and *Run*, pre-assigned in Skyline according to our experimental design (See Supplementary table 5 and Supplementary table 6). Data processing was first performed, transforming all transition intensities into $\log_2$ values. Next, normalization using global standard proteins was done on the basis of the transitions for the two housekeeping proteins (G3P and HS90A). This equalized the endogeneous intensities of global standard proteins (i.e. house keeping proteins) across all MS runs and applied similar between-run shifts to the remaining endogeneous proteins in the experiment. Quantification of protein abundance level and testing for differential abundance in the patient groups were performed with the linear mixed effect model for SRM workflows utilising heavy isotope-labelled reference peptides (Chang et al., 2012; Choi et al., 2014). Calculated P values were adjusted to control the FDR at a cut-off of 0.02. Proteins with a P value below 0.02 and a fold change larger than 1.2 were considered to be significantly dysregulated in adenoma and/or adenocarcinoma in comparison to the normal mucosa.

5.2.3 RESULTS

5.2.3.1 Development of SRM assays for candidate colorectal tumor markers

Our objective was to create a quantitative workflow to detect and measure the abundance of tumor biomarkers in colorectal adenomas and carcinomas in order to 1) verify tumor markers from our previous shotgun proteomics analysis (Uzozie et al., 2014), and 2) to establish SRM assays that would be applied on blood samples for the diagnosis of these localized colorectal neoplasms. The features of the neoplastic lesions are listed in Table 5. Forty target proteins were selected (Table 6). This comprised [1] twenty-nine candidates reported in our previous discovery phase study to be differentially expressed in colorectal precancerous lesions (Uzozie et al., 2014); [2] three candidates with increased gene expression in adenomas (Cattaneo et al., 2011), [3] two global standard control proteins, and [4] six proteins proposed to be involved in the same or related pathways as [1] (Table 6). We included adenomas as well as carcinomas since most of the candidate markers we identified in the former (Uzozie et al., 2014) have been found dysregulated also in colorectal cancers by others (Besson et al., 2011; de Wit et al., 2012; F Lam et al., 2010; Uzozie et al., 2014). Furthermore, since the chance of finding blood biomarkers of preinvasive adenomas might be lower than that for carcinomas, we intended to test our targets in both types of tissue.

Targeted SRM assays for the 40 selected proteins were developed and optimized as illustrated in Figure 14. Most proteotypic peptides for our target proteins were identified from an initial study where sample fractionation was performed prior to shotgun analysis (Uzozie et al., 2014), and from accessible data in Peptide Atlas. We also performed label-free shotgun analysis on unfractionated samples from 5 randomly selected colorectal tissues pairs. This was essential to determine the detectability of these endogenous peptides in the complex matrix of our unfractionated samples. Although sample fractionation facilitates improved sensitivity when combined with discovery proteomic workflows, it is not fully compatible with SRM workflows (Hüttenhain et al., 2009). This is because additional sample preparation steps can introduce variation in abundance measurement. Moreover, the level of sample throughput is reduced since the different fractions from each sample have to be measured individually. The 140 peptides chosen for our target 40 proteins had the best mass spectrometry performance and are listed in Table 6.

With the aid of heavy isotope-labelled reference peptides, we created SRM assays for 140 proteotypic peptides. Reference peptides with corresponding amino acid sequence were synthesized for each endogenous peptide. We generated a spectral library with data acquired after shotgun analysis on a mixture comprising all heavy reference peptides. At least 4 to 8 transitions per peptide were initially tested in an unscheduled analysis on the reference peptide pool. This resulted in peptide transitions with optimum mass spectrometry performance and iRT values (Escher et al., 2012) for defining scheduled methods to monitor our target proteins. The use of heavy isotope-labelled peptides also facilitated the elimination of interfering transitions from the endogeneous peptides. We further optimized relevant SRM parameters including collision energy, reference peptide spike-in concentration, and dwell time for each transition. We successfully created SRM assays for 74 peptides, corresponding to 33 proteins and applied these assays to detect and quantify protein abundance in tissue samples.

5.2.3.2 Targeted SRM measurement on colorectal tissues

An obvious benefit of SRM technique is the ability to perform sensitive and reproducible quantitative measurement on several target analytes in multiple samples. We utilized our SRM assays to monitor 33 target proteins in 72 colorectal tissues samples (Figure 15). The same amount of each heavy peptide was introduced into each sample from a pooled reference peptide mixture. In addition to this, iRT peptides (Escher et al., 2012) were also added to each tissue peptide sample. Heavy reference peptides were used as internal controls during statistical analysis of our SRM data, while iRT reference peptides served as internal controls to monitor instrument and technical variations during scheduled SRM measurements. Based on precursor and fragment ion intensities, at least two to four transitions were selected for each peptide. To maintain a minimum dwell time of 0.020 s per transition along the elution gradient, we split the 600 transitions in two scheduled methods based on protein number (Figure 15). Only transitions for iRT peptides and global standard proteins were analyzed in both methods. The total number of transitions and all SRM parameters for each scheduled method is listed in Supplementary tables 1 and 2 respectively.

5.2.3.3 Reproducibility of SRM measurement

Two main categories of variability can affect the sensitivity and validity of SRM measurements (Parker and Borchers, 2014). Firstly, pre-analytical variability can occur from sources such as sample preparation. And analytical variability could result from

undesirable variation in instrument parameters during SRM data acquisition. Our block-randomized study design (Oberg and Vitek, 2009) ensured complete randomization and balance in sample allocation at the sample processing and SRM spectral acquisition phases respectively (Figure 15). Protein extracts from a total of 72 tissue samples were obtained in a manner that involved the random processing of the same number of adenoma and adenocarcinoma pairs per experiment. Spectral acquisition for samples pairs was also randomized as described in *Experimental Procedure*. The paired feature for each patient was maintained in our randomized design. In addition to controlling for bias due to potentially unknown experimental artifacts during sample processing, we minimized confounding effects from instrument variance on our measured data and also on the inferred results. We examined our SRM data for unwanted bias due to variation in instrument performance.

First, the measured retention time for each of the 11 iRT peptides monitored in the sample background was compared across all the SRM runs. As shown in Supplementary figure 7, a retention time shift of less than 1 min was maintained for each iRT peptide in all samples analyzed with scheduled SRM. Even though each sample was analyzed twice with the two scheduled methods, instrument parameters were stable throughout the SRM measurements.

Next, we investigated the reproducibility of our SRM spectral peak measurements for each peptide by comparing the endogenous light to heavy peptide (L/H) peak ratios for each peptide monitored in the duplicate digest samples (Supplementary figure 8). Three pairs of tumor/normal samples (6 samples) were digested twice (Figure 15) and each of the 12 digest samples were analysed separately, according to a randomized SRM sequence. For each duplicate digest pair we compared, an R^2 value greater than 0.9 and intercept of approximately 1 was observed, indicating that a highly reproducible peak area was achieved for all peptides monitored in each duplicate pair.

5.2.3.4 Identification of peptide peak groups and corresponding proteins

SRM data acquisition on the triple quadrupole instrument was constantly monitored in SpectroDive and Skyline. We achieved automated, high-confidence identification of peptides using the mProphet statistical tool, implemented in SpectroDive. mProphet employs a 'decoy-transition' approach to probabilistically score individual features in the SRM peaks and integrates them into a combined discriminant score (Cscore) while controlling for the false discovery rate (FDR) (Reiter et al., 2011). Decoy transitions are transitions for peptide species that do no exist in the biological sample and they function as negative controls. We opted for the synthetic decoy

approach because our workflow included heavy isotope-labelled internal standards. SRM spectral data files for each sample, and a list of defined parameters for the 74 endogenous and reference peptides were imported into the SpectroDive user workspace. A comprehensive list of the peptide parameters is detailed in Supplementary table 3. For further processing, we selected peak groups with a Cscore threshold greater than 10 at a controlled FDR (Qvalue) cutoff of less than 1%. At this stringent cutoff, we achieved a sensitivity of 71.48% in confident peak group selection. Additionally, 28 target proteins were confidently identified with the 53 peptides that satisfied these criteria.

5.2.3.5 Quantification of protein abundance in tissue samples

The 53 high-confidence peptide peak groups were initially visualized in Skyline using the acquired SRM spectra file. This was necessary to confirm the absence of incorrect transitions. More so, a peptide peak group was chosen for protein quantification only if the reference groups were observed in each of the 72 samples analyzed, and the endogenous groups were detected in at least 6 samples. In addition to this, peak groups for the global standard proteins were selected only if both reference and endogenous peptides were seen in all samples. The 39 peptides that fulfilled these conditions were used to quantify a total of 25 of our protein targets. Therefore, we achieved a success rate of 62.5% (that is 25 out of 40) in the development of SRM assays to reproducibly and consistently quantify selected proteins in colorectal tissues **(**Figure 16**).**

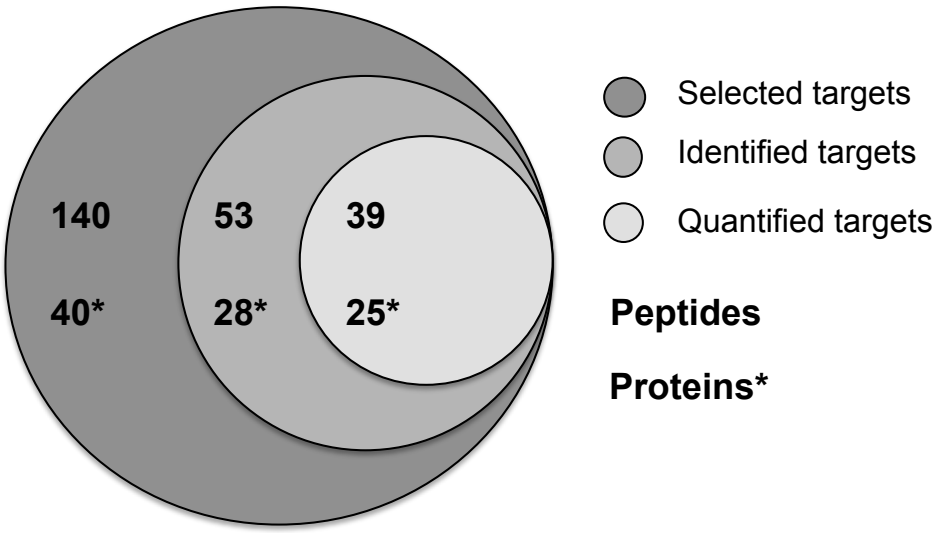


Figure 16. Target proteins and peptides selected, identified, and quantified in colorectal tissues

Each shaded circle represents the number of candidate tumor peptides and corresponding proteins selected for targeted proteomics, the total number identified with our SRM workflow, and the number of proteins quantified in normal and neoplastic lesions. Protein number is indicated with *.

Each protein was quantified with one to three proteotypic peptides (Supplementary figure 9)**.** We proceeded with statistical analysis on proteins quantified with just one peptide because the proteotypic peptide verified protein identity. Also, the corresponding heavy reference peptide was included in all samples to serve as internal controls and as a standard reference for the normalization of the endogenous peptide intensity.

5.2.3.6 Verification of putative biomarkers for colorectal neoplasms by SRM

We determined the capacity of the SRM assays developed in this study to monitor candidate biomarkers for colorectal cancer in a relatively large group of patient tumors. Tumors were stratified into two categories: the early-stage precancerous adenomas, and the invasive cancerous adenocarcinomas. With our tests, we demonstrate that 25 of our candidates are detectable by SRM in unfractionated digests of the tissue proteome. We proceeded to confirm that the SRM assays we developed could reproducibly quantify the proteins across patient tumor samples and detect protein expression changes in tumor and normal mucosa samples. In total, we monitored 25 proteins across 72 tissues samples - 19 adenomatous and matching normal mucosa samples; 17 adencarcinomas and paired samples of the normal mucosa. We confidently identified peptides and proteins detected in the sample groups with

mProphet (Reiter et al., 2011) (implemented in SpectroDive) and performed protein significance analysis with MSstats (Chang et al., 2012; Choi et al., 2014).

Proteins with an adjusted P value below 0.02 and a fold change greater then 1.2 were considered significant. Eighteen proteins showed significant abundance differences comparing colorectal adenomas and normal mucosa samples (Figure 17, and **Table 7**). The expression levels of ALDR, APEX1, ARF6, KCD12 and PGM5 were significantly lower, while proteins AN32A, ANXA3, DHSO, G6PD, LDHA, NGAL, NUCL, S10AB, SPB5, CDV3, IPYR, OLFM4, REG4 were significantly over expressed. A total of 17 proteins displayed adenocarcinoma-related protein expression changes. The expression of proteins ALDR ARF6, KCD12 and PGM5 were significantly downregulated, while that of AN32A, ANXA3, DHSO, G6PD, LDHA, NGAL, NUCL, S10AB, SPB5, REG4, AN32B, RCN1 and SET were significantly upregulated (Figure 17, and **Table 7**).

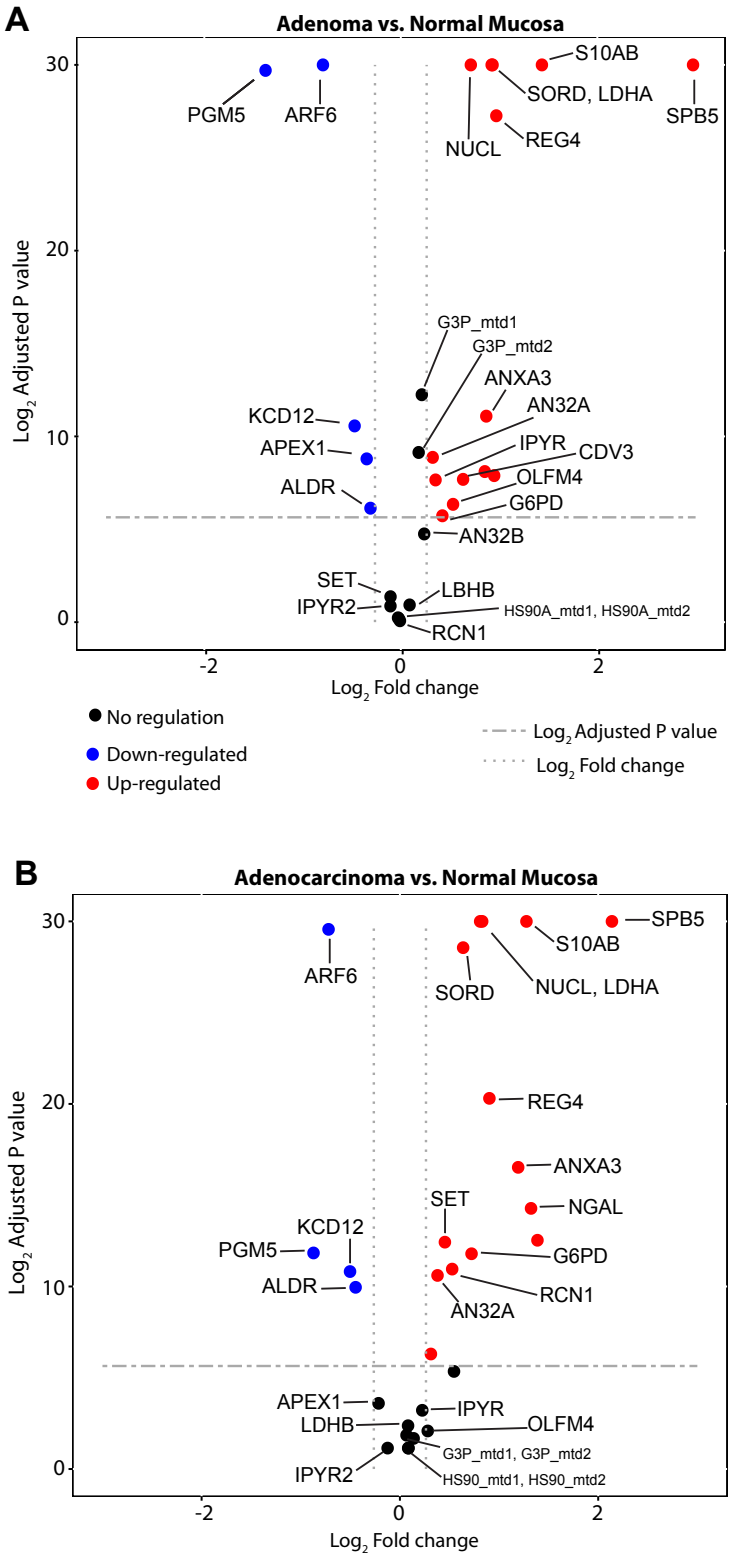


Figure 17. Quantification of target proteins in colorectal neoplasms

Protein abundance differences in adenomas (A) and adenocarcinomas (B). Proteins with an adjusted P value below 0.02, and a fold change above 1.2 were considered significant. Significantly dysregulatedproteins in tumor tissues compared to the matching sample of the normal mucosa are highlighted in red (upregulated) and blue (downregulated) dots respectively.

Table 7. SRM-based quantification of target proteins in colorectal adenomas and adenocarcinomas.

		Adenoma vs. Normal mucosa				Adenocarcinoma vs. Normal mucosa		
Protein name	**Gene name**	**Fold change**	**Adjusted P value**	**Significant**	**Trend reported in Uzozie et al.**	**Fold change**	**Adjusted P value**	**Significant**
ALDR	*AKR1B1*	1.24	0.01422872	Yes ↓	Down	1.36	0.00100745	Yes ↓
AN32A	*ANP32A*	1.25	0.00213412	Yes ↑	Up	1.30	0.00064141	Yes ↑
ANXA3	*ANXA3*	1.93	0.00421451	Yes ↑	Up	2.62	0.00016874	Yes ↑
APEX1	*APEX1*	1.27	0.00225812	Yes ↓	Up	1.16	0.08251550	No
ARF6	*ARF6*	1.73	0.00000000	Yes ↓	Up	1.65	0.00000000	Yes ↓
DHSO	*SORD*	1.91	0.00000000	Yes ↑	Up	1.56	0.00000000	Yes ↑
G3P_mtd1	*GAPDH*	1.16	0.00020606	No	NA	1.05	0.27636653	No
G6PD	*G6PD*	1.34	0.01887495	Yes ↑	NA	1.65	0.00028178	Yes ↑
HS90A_mtd1	*HS90AA1*	1.01	0.87666686	No	NA	1.06	0.45145255	No
LDHA	*LDHA*	1.90	0.00000000	Yes ↑	Up	1.78	0.00000000	Yes ↑
LDHB	*LDHB*	1.06	0.52593138	No	Up *	1.10	0.31248696	No
NGAL	*LCN2*	1.81	0.00362583	Yes ↑	Up	2.50	0.00005031	Yes ↑
NUCL	*NCL*	1.64	0.00000000	Yes ↑	Up	1.75	0.00000000	Yes ↑
S10AB	*S100A11*	2.71	0.00000000	Yes ↑	Up	2.42	0.00000000	Yes ↑
SPB5	*SERPINB5*	7.86	0.00000000	Yes ↑	Up	4.40	0.00000000	Yes ↑
AN32B	*ANP32B*	1.18	0.03715907	No	Up *	1.24	0.01267792	Yes ↑
CDV3	*CDV3*	1.55	0.00485012	Yes ↑	Up	1.46	0.02455417	No
G3P_mtd2	*GAPDH*	1.13	0.00177703	No	NA	1.06	0.19298791	No
HS90A_mtd2	*HS90AA1*	1.02	0.85064932	No	NA	1.06	0.45145255	No
IPYR	*PPA1*	1.28	0.00495606	Yes ↑	Up *	1.17	0.10750496	No
IPYR2	*PPA2*	1.08	0.54662624	No	Up *	1.09	0.45145255	No
KCD12	*KCTD12*	1.39	0.00065986	Yes ↓	Down	1.42	0.00055367	Yes ↓
OLFM4	*OLFM4*	1.45	0.01231441	Yes ↑	Up	1.21	0.23544211	No
PGM5	*PGM5*	2.60	0.00000000	Yes ↓	Down	1.83	0.00027326	Yes ↓
RCN1	*RCN1*	1.01	0.94704940	No	Up	1.44	0.00050460	Yes ↑
REG4	*REG4*	1.96	0.00000001	Yes ↑	Up	1.87	0.00000077	Yes ↑
SET	*SET*	1.08	0.38705860	No	Up	1.37	0.00018142	Yes ↑

Proteins upregulated (↑) or downregulated (↓) in colorectal adenomas and/or adenocarcinomas compared to the normal mucosa. Significantly dysregulated proteins met the cutoff criteria: fold change >1.2, adjusted P value <0.02. The global standard proteins were monitored in the two scheduled SRM runs (denoted as _mtd1 and _mtd2) and they consistently show no significant changes in expression level. NA indicates proteins that were not selected from Uzozie et al., while indistinguishable isoforms reported in Uzozie et al., are denoted with *.

Fourteen (14) proteins displaying significant tumor-related dysregulation had very similar fold change pattern in both adenomas and adenocarcinomas (**Table 7** and Figure 18). Ten of these proteins (AN32A, ANXA3, SORD, G6PD, LDHA, NGAL, NUCL, S10AB, SPB5, and REG4) were significantly upregulated in both neoplastic lesions compared to the normal mucosa.

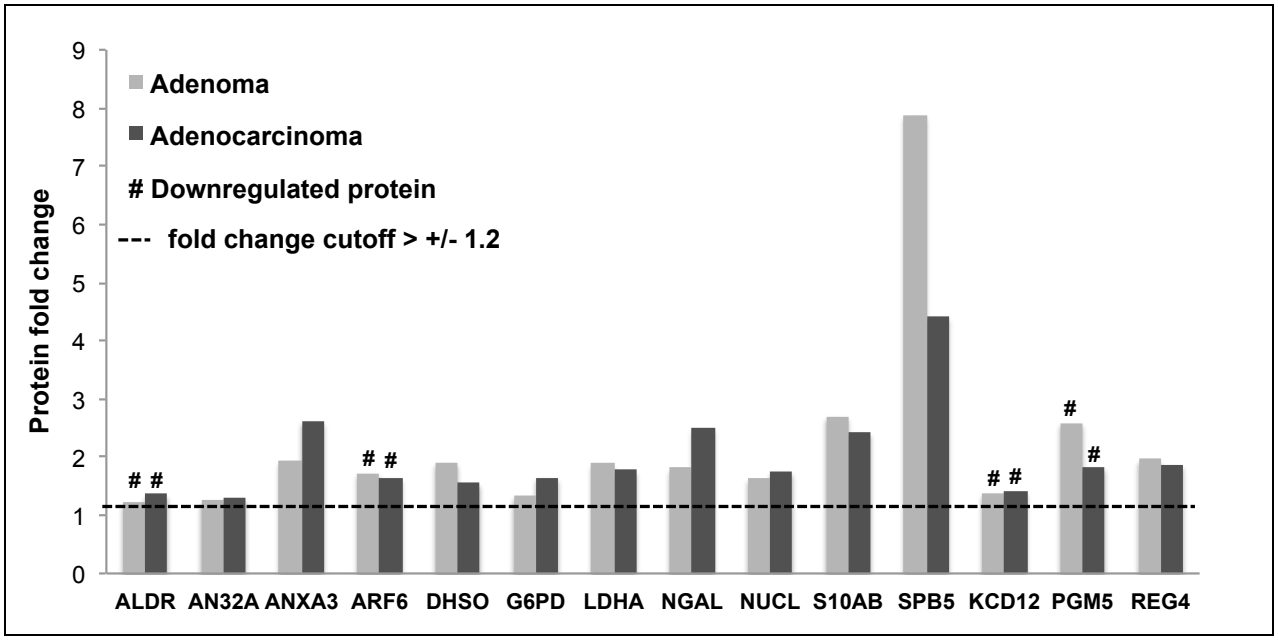


Figure 18. Fold changes for candidate proteins in adenomas and adenocarcinomas

Fourteen proteins displayed similar direction of fold change in adenomas and adenocarcinomas. Fold change cutoff is > +/-1.2 at an adjusted P value < 0.02. Downregulated proteins are denoted with #.

Excluding G6PD, all quantified proteins were targets from our previous shotgun study. Our inability to detect the three proteins (AJUBA, K1199, CADH3) selected based on data from a previous gene expression study (Cattaneo et al., 2011) and some of the other protein targets could be due to a number of factors. With SRM, the limit of detection for low abundance proteins in an unfractionated, complex sample background is most often not attained (Kulasingam et al., 2010; Surinova et al., 2011). Tissue protein concentration covers a dynamic range over 5 orders of magnitude and we did not include additional enrichment or sample fractionation steps to improve the detection of low abundance proteins. Furthermore, the endogenous proteotypic peptides targeted may be absent in our sample digests. Our protein extraction process may have eluded proteins enriched in a specific cellular compartment.

5.2.4 DISCUSSION

This study demonstrates the applicability of SRM technique to quantify putative biomarkers for colorectal tumors across multiple patient samples. Over the last decade, there have been several proteomic studies on the verification of multiple putative biomarkers in colorectal cancer (Fijneman et al., 2012; Hamelin et al., 2011; Han et al., 2011; Kume et al., 2014; Murakoshi et al., 2011; Xie et al., 2010; Xue et al., 2010) but none primarily focused on early stage tumors (de Wit et al., 2012; Jimenez et al., 2010). Our approach to investigate both colorectal adenomas and adenocarcinomas provided a consistent means to detect protein changes that occur early in colorectal tumorigenesis and that are likely to persist during the transformation of adenomas to cancerous lesions.

Literature is vast with numerous candidate biomarkers for colorectal cancer but till date, very few of these proteins have been further verified for clinical application (Luo et al., 2013; Shi et al., 2011). This is largely due to a lack of robust methods for the reproducible verification of candidate tumor markers in multiple samples. Recent developments in SRM-based targeted proteomics strongly affirm its capacity to complement conventional immunoassay-based techniques and accelerate the verification of putative biomarkers in a large number of patient samples (Maiolica et al., 2012; Picotti and Aebersold, 2012; Surinova et al., 2013). In addition to SRM (Lange et al., 2008; Picotti et al., 2009; Picotti and Aebersold, 2012), Sequential Window Acquisition of All Theoretical fragment ion spectra (SWATH) mass spectrometry (Gillet et al., 2012; Liu et al., 2013a; 2013b) and Parallel Reaction Monitoring (PRM) (Peterson et al., 2012) are two recently reported methods with worthy future prospects for their application in clinical biomarker discovery and targeted therapeutics.

The application of proteomics in biomarker research generally involves two main phases: a discovery phase to identify promising target analytes and a verification/validation phase to determine the suitability of these analytes as markers for the tumor(s) of interest in large scale studies. Indeed studies on colorectal cancer have yielded potential biomarker candidates (de Wit et al., 2012; Shi et al., 2011) from discovery stage studies employing gel electrophoresis and more recently, shotgun proteomics (Albrethsen et al., 2010; Besson et al., 2011; F Lam et al., 2010). But most studies fall short of verifying identified putative markers to advance their clinical application. It has been shown that most research has focused on only a few extensively studied proteins solely because the analytical tools for their study are available (Edwards et al., 2011).

We developed SRM assays for 40 proteins (Table 6), most of which have been previously reported in a discovery phase study (Uzozie et al., 2014). Without sample fractionation and depletion of high abundance proteins we verified promising tumor markers for colorectal tumorigenesis by SRM-based quantification of proteins in precancerous and cancerous lesions. Protein abundance measurements were achieved with the use of heavy isotope-labelled reference standards, and recently updated tools (mProphet, Skyline and MSstats) for sensitive and reproducible SRM analysis. SRM assays were created in a multistep approach (Figure 14 and Figure 15), in which SRM coordinates were first obtained from fragment spectra with heavy reference peptides, refined by measuring these peptides in scheduled SRM mode, and built into definitive assays for peptide and protein quantification. The positive evaluation of the sample preparation protocol and SRM measurements in terms of sensitivity and reproducibility (Supplementary figure 7 and Supplementary figure 8) demonstrate that optimized SRM assays can be accurately replicated and hence adapted for use in various laboratories.

Among the 40 proteins we developed SRM assays for, 28 were confidently detected in colorectal adenomas with stringent peptide scoring thresholds (Cscore greater than 10, FDR less than 1%). Most of these proteins had been observed in extensively fractionated colorectal tissue samples by mass spectrometry (Uzozie et al., 2014). It was therefore anticipated that some proteins would not be detected in the unfractionated, complex samples for SRM analysis. As shown in Figure 16, 62.5% of the selected proteins (25 of 40) were consistently quantified across tissue samples. The direction of abundance changes for most proteins with tumor-related dysregulation (Figure 17 and **Table 7**) was consistent with the discovery phase results that were based on iTRAQ 8plex shotgun proteomics (Uzozie et al., 2014). Literature review suggests that none of the proteins we detected with significant abundance changes in colorectal neoplasms (Figure 17 and Figure 18) have been validated in clinical biomarker studies (de Wit et al., 2012; Jimenez et al., 2010; Kim, 2009; Luo et al., 2013). Based on discovery proteomic studies performed on clinical material (human tissue, plasma, serum, and faeces) from colorectal cancer patients, a number of these proteins have been reported to show differential expression levels in carcinomas when compared to the normal mucosa (Besson et al., 2011; Conrotto et al., 2008; de Wit et al., 2012; F Lam et al., 2010; Friedman et al., 2004; Jimenez et al., 2010; Kim et al., 2006; Luo et al., 2013; Shi et al., 2011; Uzozie et al., 2014; Watanabe et al., 2008).

Our SRM analysis identified 10 proteins, consisting of AN32A, ANXA3, SORD, G6PD, LDHA, NGAL, NUCL, S10AB, SPB5, and REG4, which were significantly

upregulated in colorectal adenomas and adenocarcinomas compared to the normal mucosa (Figure 17 and Figure 18). Even though proteins SET and RCN1 were not significantly overexpressed in adenomas, adenocarcinomas displayed significant elevated levels for both proteins in comparison to normal mucosa samples.

Examination of documented reports indicates that the 10 proteins with significant abundance changes in early and late neoplastic lesions of the large bowel are involved in processes that could contribute to tumorigenesis and progression of cancer. And three of them (LDHA, SORD, G6PD) are indicative of metabolic reprogramming to enable robust biosynthesis and antioxidant defence. **SET** and **AN32A** are often co-isolated and are reported to be parts of a protein complex involved in chromatin remodelling through the inhibition of histone acetylation (Seo et al., 2001; Shi et al., 2014). Depending on the nature of intracellular signals, SET and AN32A containing complex may either stimulate or silence gene transcriptions (Cervoni, 2002; Seo et al., 2002; Shi et al., 2014). Regarding cell proliferation, SET could either positively regulate cell cycle by activating cyclin E/cdk2 or suppress cell proliferation by inhibiting the ERK pathway (Estanyol et al.; Fukukawa et al., 2005; Shi et al., 2014). The overexpression of **LDHA** in adenomas and adenocarcinomas is representative of increased pyruvate metabolism, which is further associated with increased glucose metabolism via aerobic glycolysis. This established alteration in metabolism is further supported by discoveries that cancer-associated genes regulate various enzymes involved in glucose breakdown (Hsu and Sabatini, 2008; Koppenol et al., 2011; Sillars-Hardebol et al., 2012; WARBURG, 1956; Ward, 2012). For example LDHA is activated by HIF1 (Koppenol et al., 2011). SORD is a metabolic enzyme involved in the polyol pathway. We recently reported evidence of a dysregulated polyol pathway during colorectal cancer carcinogenesis in a discovery phase study on colorectal adenomas (Uzozie et al., 2014). This SRM data also concurs with our previous report on the other enzyme of the polyol pathway, ALDR (**Table 7**), which we found to be downregulated in *adenomatuous lesions* (Uzozie et al., 2014). **G6PD** is the rate-limiting enzyme of the pentose phosphate pathway (PPP). While G6PD may be dispensable for pentose synthesis, it is crucial for cellular defence against redox-stress-induced apoptosis (Fico et al., 2004; P P Pandolfi, 1995). **NUCL** is a highly phosphorylated multi-functional protein located in the nucleolus. It's expression is linked to the growth rate of the cell, and a loss of NUCL expression has been shown to occur in differentiating cells (Bicknell et al., 2005). The potential functions of the secreted protein, **REG4**, include cell proliferation, migration, invasion, and resistance to apoptosis (Rafa et al., 2010; Violette et al., 2003). Its expression is also regulated by signalling events linked to

known tumorigenic pathways (Beelen Granlund et al., 2013; Rafa et al., 2010; Violette et al., 2003). **ANXA3** is considered to be a novel angiogenic factor that induces vascular endothelial growth factor production through the HIF-1 pathway (Mussunoor and Murray, 2008; Park et al., 2005). Furthermore, **NGAL** overexpression in neoplastic cells promotes tumor invasion while decreasing E-cadherin-mediated cell-cell adhesion (Conrotto et al., 2008; Sun et al., 2011). The function of **S10AB** (also known as S100A11) in carcinogenesis is still unclear. It is best described as a dual mediator that acts as a tumor suppressor gene in some cancers (such as breast cancer) and as an oncogenic promoter in other cancers (such as colorectal cancer), depending on its localization and interacting partners (Emberley et al., 2004; He et al., 2009). Similar to S10AB, oncogenic and tumor suppressor activities are reported for **SPB5** (Snoeren et al., 2013). Recent biomarker studies on colorectal cancer correlate elevated SPB5 expression to CEA levels and poor patient prognosis (Baek et al., 2013; Snoeren et al., 2013).

The targeted SRM workflow described here provides sensitive SRM assays to enable the direct detection and consistent quantification of putative biomarkers. Using the outlined workflow, together with the software tools mProphet, Skyline and MSstats accelerates assay development and improves the high-confidence analysis of large-scale SRM data sets. This approach is suitable for the verification of several putative biomarkers in a large group of patient tissue samples and provides a basis for further clinical verification studies on the selected biomarkers.

In line with the goals of a biomarker verification study, a worthwhile future step would be to apply the established SRM assays to profile the identified candidate biomarkers in plasma samples from patients with colorectal cancers and individuals without the disease (controls). Ideally, this should be a large cohort study, representative of the target population. We have commenced verification studies to evaluate the performance of our SRM workflow in plasma samples. It is known that plasma proteins range in concentration over 12 orders of magnitude and over 95% of the protein mass is comprised of a few high abundance proteins (Anderson, 2002; Kulasingam et al., 2010). Regardless of this issue of sample complexity, current improvements in mass spectrometry have facilitated the quantitation of low and high abundance plasma proteins (Addona et al., 2009; Anderson and Hunter, 2006; Anderson et al., 2009; 2004; Hüttenhain et al., 2009; Kuzyk et al., 2009; Surinova et al., 2011). As tissue-derived proteins are often diluted in the systematic blood stream, our SRM workflow could be refined to include affinity enrichment techniques, depletion of high abundance plasma proteins or fractionation steps to improve detection

sensitivity for low abundance proteins(Meng and Veenstra, 2011). Using plasma samples, SRM assays for each candidate biomarker would be examined in terms of dynamic range, reproducibility, limit of detection, specificity and sensitivity for the diseased state.

5.2.5 REFERENCES

Addona, T.A., Abbatiello, S.E., Schilling, B., Skates, S.J., Mani, D.R., Bunk, D.M., Spiegelman, C.H., Zimmerman, L.J., Ham, A.-J.L., Keshishian, H., et al. (2009). Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol 27, 633–U685.

Albrethsen, J., Knol, J.C., Piersma, S.R., Pham, T.V., de Wit, M., Mongera, S., Carvalho, B., Verheul, H.M.W., Fijneman, R.J.A., Meijer, G.A., et al. (2010). Subnuclear Proteomics in Colorectal Cancer. … & Cellular Proteomics.

Anderson, L., and Hunter, C.L. (2006). Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol. Cell Proteomics 5, 573–588.

Anderson, N.L. (2002). The Human Plasma Proteome: History, Character, and Diagnostic Prospects. Molecular & Cellular Proteomics 1, 845–867.

Anderson, N.L., Anderson, N.G., Pearson, T.W., Borchers, C.H., Paulovich, A.G., Patterson, S.D., Gillette, M., Aebersold, R., and Carr, S.A. (2009). A Human Proteome Detection and Quantitation Project. Molecular & Cellular Proteomics 8, 883–886.

Anderson, N.L., Polanski, M., and Pieper, R. (2004). The human plasma proteome. … Cell Proteomics.

Ang, C.-S., Phung, J., and Nice, E.C. (2011). The discovery and validation of colorectal cancer biomarkers. Biomedical Chromatography 25, 82–99.

Anon (2003). The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest. Endosc. 58, S3–S43.

Baek, J.Y., Yeo, H.Y., Chang, H.J., Kim, K.-H., Kim, S.Y., Park, J.W., Park, S.C., Choi, H.S., Kim, D.Y., and Oh, J.H. (2013). Serpin B5 is a CEA-interacting biomarker for colorectal cancer. Int. J. Cancer 134, 1595–1604.

Beelen Granlund, A., Østvik, A.E., Brenna, Ø., Torp, S.H., Gustafsson, B.I., and Sandvik, A.K. (2013). REG gene expression in inflamed and healthy colon mucosa explored by in situ hybridisation. Cell Tissue Res 1–8.

Besson, D., Pavageau, A.-H., Valo, I., Bourreau, A., Bélanger, A., Eymerit-Morin, C., Moulière, A., Chassevent, A., Boisdron-Celle, M., Morel, A., et al. (2011). A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. Mol. Cell Proteomics 10, M111.009712.

Bicknell, K., Brooks, G., Kaiser, P., Chen, H., Dove, B.K., and Hiscox, J.A. (2005). Nucleolin is regulated both at the level of transcription and translation. Biochemical and Biophysical Research Communications 332, 817–822.

Cattaneo, E., Baudis, M., Buffoli, F., Bianco, M.A., Zorzi, F., and Marra, G. (2010). Pathways and Crossroads to Colorectal Cancer. In Pre-Invasive Disease: …, (New York, NY: Springer New York), pp. 369–394.

Cattaneo, E., Laczko, E., Buffoli, F., Zorzi, F., Bianco, M.A., Menigatti, M., Bartosova, Z., Haider, R., Helmchen, B., Sabates-Bellver, J., et al. (2011). Preinvasive

colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid vs. nonpolypoid). EMBO Mol Med 3, 334–347.

Cervoni, N. (2002). The Oncoprotein Set/TAF-1beta , an Inhibitor of Histone Acetyltransferase, Inhibits Active Demethylation of DNA, Integrating DNA Methylation and Transcriptional Silencing. Journal of Biological Chemistry 277, 25026–25031.

Chang, C.-Y., Picotti, P., Hüttenhain, R., Heinzelmann-Schwarz, V., Jovanovic, M., Aebersold, R., and Vitek, O. (2012). Protein significance analysis in selected reaction monitoring (SRM) measurements. Mol. Cell Proteomics 11, M111.014662.

Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics.

Conrotto, P., Roesli, C., Rybak, J., Kischel, P., Waltregny, D., Neri, D., and Castronovo, V. (2008). Identification of new accessible tumor antigens in human colon cancer by ex vivo protein biotinylation and comparative mass spectrometry analysis. Int. J. Cancer 123, 2856–2864.

Cunningham, D., Atkin, W., Lenz, H.-J., Lynch, H.T., Minsky, B., Nordlinger, B., and Starling, N. (2010). Colorectal cancer. The Lancet 375, 1030–1047.

Davies, R.J., Miller, R., and Coleman, N. (2005). Colorectal cancer screening: prospects for molecular stool analysis. Nat Rev Cancer 5, 199–209.

de Wit, M., Fijneman, R.J.A., Verheul, H.M.W., Meijer, G.A., and Jimenez, C.R. (2012). Proteomics in colorectal cancer translational research: Biomarker discovery for clinical applications. Clin. Biochem.

Domon, B., and Aebersold, R. (2006). Mass Spectrometry and Protein Analysis. Science 312, 212–217.

Domon, B., and Aebersold, R. (2010). Options and considerations when selecting a quantitative proteomics strategy. Nat Biotechnol 28, 710–721.

Edwards, A.M., Isserlin, R., Bader, G.D., Frye, S.V., Willson, T.M., and Yu, F.H. (2011). Too many roads not taken. Nature 470, 163–165.

Emberley, E.D., Murphy, L.C., and Watson, P.H. (2004). S100 proteins and their influence on pro-survival pathways in cancer. Biochem. Cell Biol. 82, 508–515.

Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M.J., and Rinner, O. (2012). Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 12, 1111–1121.

Estanyol, J.M., Jaumot, M., Casanovas, O., Rodriguez-Vilarrupla, A., Agell, N., and Bachs, O. The Protein SET Regulates the Inhibitory Effect of p21Cip1 on Cyclin E-Cyclin-dependent Kinase 2 Activity.

F Lam, F., Jankova, L., Dent, O.F., Molloy, M.P., Kwun, S.Y., Clarke, C., Chapuis, P., Robertson, G., Beale, P., Clarke, S., et al. (2010). Identification of distinctive protein expression patterns in colorectal adenoma. Prot. Clin. Appl. 4, 60–70.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell 61, 759–767.

Fico, A., Paglialunga, F., Cigliano, L., Abrescia, P., Verde, P., Martini, G., Iaccarino, I., and Filosa, S. (2004). Glucose-6-phosphate dehydrogenase plays a crucial role in protection from redox-stress-induced apoptosis. Cell Death Differ 11, 823–831.

Fijneman, R.J.A., de Wit, M., Pourghiasian, M., Piersma, S.R., Pham, T.V., Warmoes, M.O., Lavaei, M., Piso, C., Smit, F., Delis-van Diemen, P.M., et al. (2012). Proximal fluid proteome profiling of mouse colon tumors reveals biomarkers for early diagnosis of human colorectal cancer. Clin. Cancer Res. 18, 2613–2624.

Friedman, D.B., Hill, S., Keller, J.W., Merchant, N.B., Levy, S.E., Coffey, R.J., and Caprioli, R.M. (2004). Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. Proteomics 4, 793–811.

Fukukawa, C., Shima, H., Tanuma, N., Okada, T., Kato, N., Adachi, Y., and Kikuchi, K. (2005). The oncoprotein I-2PP2A/SET negatively regulates the MEK/ERK pathway and cell proliferation. Int. J. Oncol. 26, 751–756.

Füzéry, A.K., Levin, J., Chan, M.M., and Chan, D.W. (2013). Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. Clin Proteomics 10, 13.

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Molecular & Cellular Proteomics 11, O111.016717.

Hamelin, C., Cornut, E., Poirier, F., Pons, S., Beaulieu, C., Charrier, J.-P., Haïdous, H., Cotte, E., Lambert, C., Piard, F., et al. (2011). Identification and verification of heat shock protein 60 as a potential serum marker for colorectal cancer. FEBS Journal 278, 4845–4859.

Han, C.-L., Chen, J.-S., Chan, E.-C., Wu, C.-P., Yu, K.-H., Chen, K.-T., Tsou, C.-C., Tsai, C.-F., Chien, C.-W., Kuo, Y.-B., et al. (2011). An informatics-assisted label-free approach for personalized tissue membrane proteomics: case study on colorectal cancer. Molecular & Cellular Proteomics 10, M110.003087.

He, H., Li, J., Weng, S., Li, M., and Yu, Y. (2009). S100A11: Diverse Function and Pathology Corresponding to Different Target Proteins. Cell Biochem Biophys 55, 117–126.

Hsu, P.P., and Sabatini, D.M. (2008). Cancer Cell Metabolism: Warburg and Beyond. Cell 134, 703–707.

Hüttenhain, R., Malmström, J., Picotti, P., and Aebersold, R. (2009). Perspectives of targeted mass spectrometry for protein biomarker verification. Current Opinion in Chemical Biology 13, 518–525.

Hüttenhain, R., Soste, M., Selevsek, N., Röst, H., Sethi, A., Carapito, C., Farrah, T., Deutsch, E.W., Kusebauch, U., Moritz, R.L., et al. (2012). Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. Sci Transl Med 4, 142ra94.

Issaq, H.J., and Veenstra, T.D. (2008). Would you prefer multiple reaction monitoring or antibodies with your biomarker validation?

Jimenez, C.R., Knol, J.C., Meijer, G.A., and Fijneman, R.J.A. (2010). Proteomics of colorectal cancer: Overview of discovery studies and identification of commonly

identified cancer-associated proteins and candidate CRC serum markers. Journal of Proteomics 73, 1873–1895.

Kim, H., Kang, H.J., You, K.T., Kim, S.H., Lee, K.Y., Kim, T.I., Kim, C., Song, S.Y., Kim, H.-J., Lee, C., et al. (2006). Suppression of human selenium-binding protein 1 is a late event in colorectal carcinogenesis and is associated with poor survival. Proteomics 6, 3466–3476.

Kim, K.E. (2009). Early Detection and Prevention of Colorectal Cancer (SLACK).

Koppenol, W.H., Bounds, P.L., and Dang, C.V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism.

Kulasingam, V., Pavlou, M.P., and Diamandis, E.P. (2010). Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. Nat Rev Cancer 10, 371–378.

Kume, H., Muraoka, S., Kuga, T., Adachi, J., Narumi, R., Watanabe, S., Kuwano, M., Kodera, Y., Matsushita, K., Fukuoka, J., et al. (2014). Discovery of colorectal cancer biomarker candidates by membrane proteomic analysisand subsequent verification using selected reaction monitoring and tissue microarrayanalysis. Molecular & Cellular Proteomics.

Kuzyk, M.A., Smith, D., Yang, J., Cross, T.J., Jackson, A.M., Hardie, D.B., Anderson, N.L., and Borchers, C.H. (2009). Multiple Reaction Monitoring-based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma. Mol. Cell Proteomics 8, 1860–1877.

Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. Mol. Syst. Biol. 4, 222.

Liu, Y., Hüttenhain, R., Collins, B., and Aebersold, R. (2013a). Mass spectrometric protein maps for biomarker discovery and clinical research. Expert Rev. Mol. Diagn. 13, 811–825.

Liu, Y., Hüttenhain, R., Surinova, S., Gillet, L.C.J., Mouritsen, J., Brunner, R., Navarro, P., and Aebersold, R. (2013b). Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 13, 1247–1256.

Lopez, M.F., Kuppusamy, R., Sarracino, D.A., Prakash, A., Athanas, M., Krastins, B., Rezai, T., Sutton, J.N., Peterman, S., and Nicolaides, K. (2011). Mass Spectrometric Discovery and Selective Reaction Monitoring (SRM) of Putative Protein Biomarker Candidates in First Trimester Trisomy 21 Maternal Serum. J Proteome Res 10, 133–142.

Luo, Y., Wang, L., and Wang, J. (2013). Developing proteomics-based biomarkers for colorectal neoplasms for clinical practice: Opportunities and challenges. Prot. Clin. Appl. 7, 30–41.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968.

Maiolica, A., Jünger, M.A., Ezkurdia, I., and Aebersold, R. (2012). Targeted proteome investigation via selected reaction monitoring mass spectrometry. Journal of Proteomics 75, 3495–3513.

Meng, Z., and Veenstra, T.D. (2011). Targeted mass spectrometry approaches for protein biomarker verification. Journal of Proteomics 74, 2650–2659.

Murakoshi, Y., Honda, K., Sasazuki, S., Ono, M., Negishi, A., Matsubara, J., Sakuma, T., Kuwabara, H., Nakamori, S., Sata, N., et al. (2011). Plasma biomarker discovery and validation for colorectal cancer by quantitative shotgun mass spectrometry and protein microarray. Cancer Science 102, 630–638.

Muraoka, S., Kume, H., Watanabe, S., Adachi, J., Kuwano, M., Sato, M., Kawasaki, N., Kodera, Y., Ishitobi, M., Inaji, H., et al. (2012). Strategy for SRM-based Verification of Biomarker Candidates Discovered by iTRAQ Method in Limited Breast Cancer Tissue Samples. J Proteome Res.

Mussunoor, S., and Murray, G.I. (2008). The role of annexins in tumour development and progression. J. Pathol. 216, 131–140.

Oberg, A.L., and Vitek, O. (2009). Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments. J Proteome Res 8, 2144–2156.

P P Pandolfi, F.S.R.R.P.M.F.G.L.L. (1995). Targeted disruption of the housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD): G6PD is dispensable for pentose synthesis but essential for defense against oxidative stress. The EMBO Journal 14, 5209.

Park, J.E., Lee, D.H., Lee, J.A., Park, S.G., Kim, N.-S., Park, B.C., and Cho, S. (2005). Annexin A3 is a potential angiogenic mediator. Biochemical and Biophysical Research Communications 337, 1283–1287.

Parker, C.E., and Borchers, C.H. (2014). Mass spectrometry based biomarker discovery, verification, and validation - Quality assurance and control of protein biomarker assays. Mol Oncol.

Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., and Coon, J.J. (2012). Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. Mol. Cell Proteomics 11, 1475–1488.

Picotti, P., Rinner, O., Stallmach, R., and Dautel, F. (2009). High-throughput generation of selected reaction-monitoring assays for proteins and proteomes : Abstract : Nature Methods. Nat. Methods.

Picotti, P., and Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat. Methods 9, 555–566.

Rafa, L., Dessein, A.-F., Devisme, L., Buob, D., Truant, S., Porchet, N., Huet, G., Buisine, M.-P., and Lesuffleur, T. (2010). REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. Int. J. Oncol. 36, 689–698.

Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M.O., and Aebersold, R. (2011). mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat. Methods 8, 430–435.

Rifai, N., Gillette, M.A., and Carr, S.A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 24, 971–983.

Selevsek, N., Matondo, M., Sanchez Carbayo, M., Aebersold, R., and Domon, B. (2011). Systematic quantification of peptides/proteins in urine using selected reaction monitoring. Proteomics 11, 1135–1147.

Seo, S.-B., Macfarlan, T., McNamara, P., Hong, R., Mukai, Y., Heo, S., and Chakravarti, D. (2002). Regulation of Histone Acetylation and Transcription by Nuclear Protein pp32, a Subunit of the INHAT Complex. Journal of Biological ….

Seo, S.-B., McNamara, P., Heo, S., Turner, A., Lane, W.S., and Chakravarti, D. (2001). Regulation of Histone Acetylation and Transcription by INHAT, a Human Cellular Complex Containing the Set Oncoprotein. Cell 104, 119–130.

Shi, H., Hood, K.A., Hayes, M.T., and Stubbs, R.S. (2011). Proteomic analysis of advanced colorectal cancer by laser capture microdissection and two-dimensional difference gel electrophoresis. Journal of Proteomics 75, 339–351.

Shi, T., Gao, Y., Quek, S.I., Fillmore, T.L., Nicora, C.D., Su, D., Zhao, R., Kagan, J., Srivastava, S., Rodland, K.D., et al. (2014). A Highly Sensitive Targeted Mass Spectrometric Assay for Quantification of AGR2 Protein in Human Urine and Serum. J Proteome Res 13, 875–882.

Siegel, R., Desantis, C., and Jemal, A. (2014). Colorectal cancer statistics, 2014. CA: a Cancer Journal for Clinicians 64, 104–117.

Siegel, R., Naishadham, D., and Jemal, A. (2013). Cancer statistics, 2013. CA: a Cancer Journal for Clinicians 63, 11–30.

Sillars-Hardebol, A.H., Carvalho, B., van Engeland, M., Fijneman, R.J.A., and Meijer, G.A. (2012). The adenoma hunt in colorectal cancer screening: defining the target. J. Pathol. 226, 1–6.

Snoeren, N., Emmink, B.L., Koerkamp, M.J.G., van Hooff, S.R., Goos, J.A.C.M., van Houdt, W.J., de Wit, M., Prins, A.M., Piersma, S.R., Pham, T.V., et al. (2013). Maspin is a marker for early recurrence in primary stage III and IV colorectal cancer. Br. J. Cancer 109, 1636–1647.

Song, K., Fendrick, A.M., and Ladabaum, U. (2004). Fecal DNA testing compared with conventional colorectal cancer screening methods: a decision analysis. Gastroenterology 126, 1270–1279.

Sun, Y., Yokoi, K., Li, H., Gao, J., Hu, L., Liu, B., Chen, K., Hamilton, S.R., Fan, D., Sun, B., et al. (2011). NGAL expression is elevated in both colorectal adenoma-carcinoma sequence and cancer progression and enhances tumorigenesis in xenograft mouse models. Clin. Cancer Res. 17, 4331–4340.

Surinova, S., Hüttenhain, R., Chang, C.-Y., Espona, L., Vitek, O., and Aebersold, R. (2013). Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. Nat Protoc 8, 1602–1619.

Surinova, S., Schiess, R., Hüttenhain, R., Cerciello, F., Wollscheid, B., and Aebersold, R. (2011). On the Development of Plasma Protein Biomarkers. J Proteome Res 10, 5–16.

Uzozie, A., Nanni, P., Staiano, T., Grossmann, J., Barkow-Oesterreicher, S., Shay, J.W., Tiwari, A., Buffoli, F., Laczko, E., and Marra, G. (2014). Sorbitol dehydrogenase overexpression and other aspects of dysregulated protein expression in human precancerous colorectal neoplasms: a quantitative proteomics study. Molecular & Cellular Proteomics 13, 1198–1218.

Violette, S., Festor, E., Pandrea-Vasile, I., Mitchell, V., Adida, C., Dussaulx, E., Lacorte, J.M., Chambaz, J., Lacasa, M., and Lesuffleur, T. (2003). Reg IV, a new

member of the regenerating gene family, is overexpressed in colorectal carcinomas. Int. J. Cancer 103, 185–193.

WARBURG, O. (1956). On the origin of cancer cells. Science 123, 309–314.

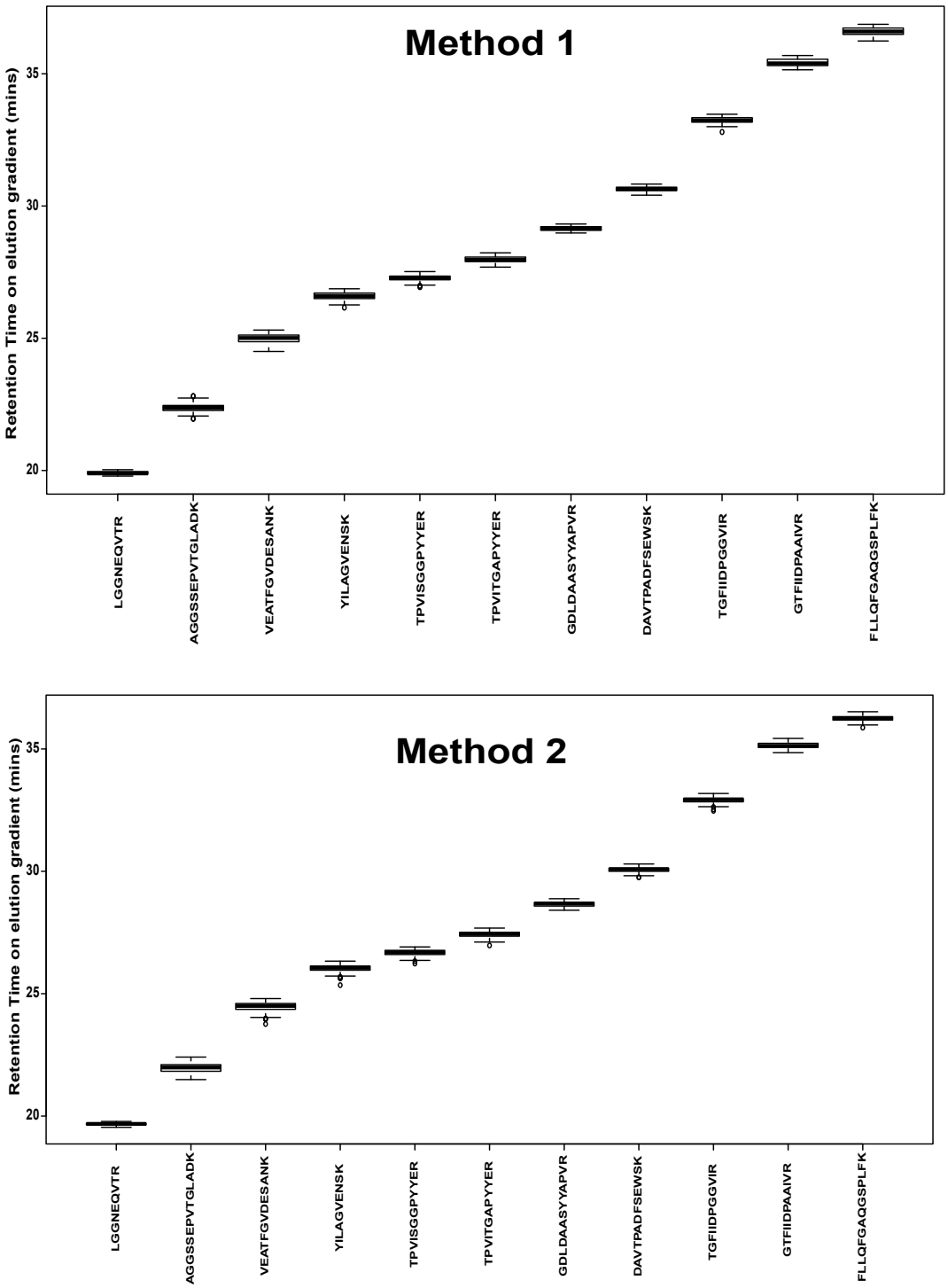
Ward, P. (2012). ScienceDirect.com - Cancer Cell - Metabolic Reprogramming: A Cancer Hallmark Even Warburg Did Not Anticipate. Cancer Cell.

Watanabe, M., Takemasa, I., Kawaguchi, N., Miyake, M., Nishimura, N., Matsubara, T., Matsuo, E.-I., Sekimoto, M., Nagai, K., Matsuura, N., et al. (2008). An application of the 2-nitrobenzenesulfenyl method to proteomic profiling of human colorectal carcinoma: A novel approach for biomarker discovery. Prot. Clin. Appl. 2, 925–935.

Xie, L.-Q., Zhao, C., Cai, S.-J., Xu, Y., Huang, L.-Y., Bian, J.-S., Shen, C.-P., Lu, H.-J., and Yang, P.-Y. (2010). Novel Proteomic Strategy Reveal Combined α 1Antitrypsin and Cathepsin D as Biomarkers for Colorectal Cancer Early Screening. J Proteome Res 9, 4701–4709.

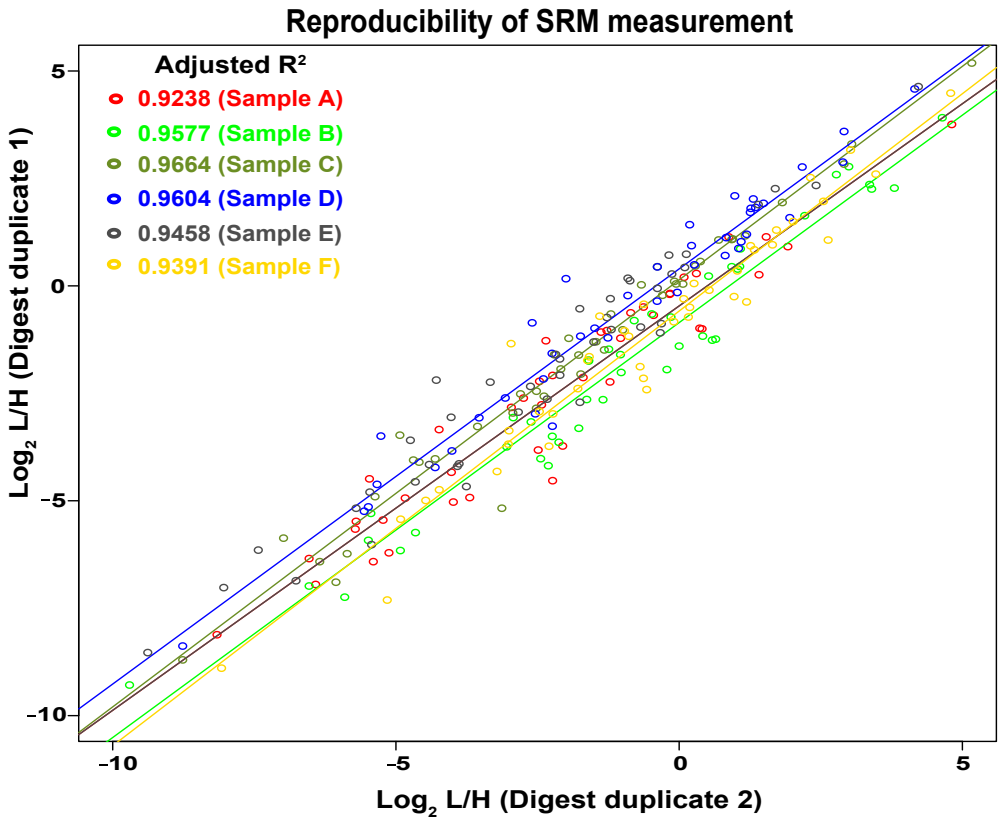
Xue, H., Lü, B., Zhang, J., Wu, M., Huang, Q., Wu, Q., Sheng, H., Wu, D., Hu, J., and Lai, M. (2010). Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. J Proteome Res 9, 545–555.

5.2.6 SUPPLEMENTARY FIGURES



Supplementary figure 7. Distribution of retention time (RT) for iRT peptides in samples.

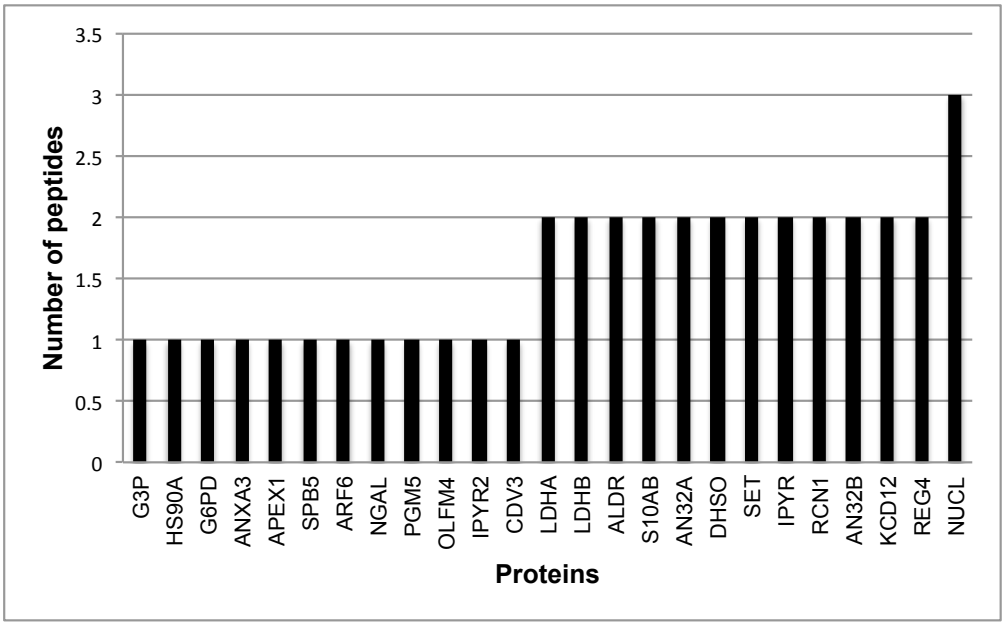
A retention time shift of <1min was observed for each of the 11 iRT peptides, measured in the sample matrix, during the scheduled SRM runs with method 1 and method 2 transitions respectively. The y-axis represents the segment of the elution gradient (20 - 35 min) within which the iRT peptides eluted. Error bars indicate the average RT of the peptide in all samples and the1st and 3rd quartiles.



	Intercept	Slope	Adjusted R-squared	p-value
Sample A (Digest duplicate 1 vs. Digest Duplicate 2)	-0.46661	0.941	0.9238	< 2.2e-16
Sample B (Digest duplicate 1 vs. Digest Duplicate 2)	-0.85477	0.96551	0.9577	< 2.2e-16
Sample C (Digest duplicate 1 vs. Digest Duplicate 2)	0.14011	0.99369	0.9664	< 2.2e-16
Sample D (Digest duplicate 1 vs. Digest Duplicate 2)	0.39564	0.96607	0.9604	< 2.2e-16
Sample E (Digest duplicate 1 vs. Digest Duplicate 2)	0.27327	0.9626	0.9458	< 2.2e-16
Sample F (Digest duplicate 1 vs. Digest Duplicate 2)	-0.58602	1.01149	0.9391	< 2.2e-16

Supplementary figure 8. Reproducibility of SRM measurements in duplicate sample digests

Six samples (3 normal and disease pairs) were digested in duplicate and analyzed during scheduled SRM experiments. Peptides monitored in method 1 and method 2 were combined for this plot. For each duplicate digest compared, the correlation of $\log_2$ light/heavy ratio for the 39 peptides used for protein quantification is shown. The intercept, slope and adjusted R^2 values for each regression are listed.



Supplementary figure 9. Quantified peptides for each protein

The abundance of twenty five proteins was quantified in colorectal tissues with a minimum of one to three high-confidence proteotypic peptide per protein (see *Results*).

5.2.7 SUPPLEMENTARY TABLES

Supplementary table 1. List of transitions selected for scheduled SRM on the TSQ Vantage (Method 1).

Q1	Q3	CE	RT-2.5	RT+2.5	Polarity	Trigger	Reference	Peptide Sequence
560.32	763.47	19.7	31.25	36.25	1	5000	0	DQLIYNLLK
560.32	650.39	19.7	31.25	36.25	1	5000	0	DQLIYNLLK
560.32	487.32	19.7	31.25	36.25	1	5000	0	DQLIYNLLK
564.33	771.49	19.7	31.25	36.25	1	5000	0	DQLIYNLLK
564.33	658.40	19.7	31.25	36.25	1	5000	0	DQLIYNLLK
564.33	495.34	19.7	31.25	36.25	1	5000	0	DQLIYNLLK
829.43	1001.55	27.8	35.33	40.33	1	5000	0	DLADELALVDVIEDK
829.43	930.51	27.8	35.33	40.33	1	5000	0	DLADELALVDVIEDK
829.43	718.36	27.8	35.33	40.33	1	5000	0	DLADELALVDVIEDK
833.44	1009.57	27.8	35.33	40.33	1	5000	0	DLADELALVDVIEDK
833.44	938.53	27.8	35.33	40.33	1	5000	0	DLADELALVDVIEDK
833.44	726.38	27.8	35.33	40.33	1	5000	0	DLADELALVDVIEDK
632.84	1037.55	21.9	25.6	30.6	1	5000	0	QVVESAYEVIK
632.84	938.48	21.9	25.6	30.6	1	5000	0	QVVESAYEVIK
632.84	809.44	21.9	25.6	30.6	1	5000	0	QVVESAYEVIK
636.85	1045.57	21.9	25.6	30.6	1	5000	0	QVVESAYEVIK
636.85	946.50	21.9	25.6	30.6	1	5000	0	QVVESAYEVIK
636.85	817.45	21.9	25.6	30.6	1	5000	0	QVVESAYEVIK
706.40	1042.59	24.1	26.81	31.81	1	5000	0	GALQNIIPASTGAAK
706.40	928.55	24.1	26.81	31.81	1	5000	0	GALQNIIPASTGAAK
706.40	815.46	24.1	26.81	31.81	1	5000	0	GALQNIIPASTGAAK
710.41	1050.60	24.1	26.81	31.81	1	5000	0	GALQNIIPASTGAAK
710.41	936.56	24.1	26.81	31.81	1	5000	0	GALQNIIPASTGAAK
710.41	823.48	24.1	26.81	31.81	1	5000	0	GALQNIIPASTGAAK
882.40	1101.46	29.4	29.99	34.99	1	5000	0	LISWYDNEFGYSNR
882.40	986.43	29.4	29.99	34.99	1	5000	0	LISWYDNEFGYSNR
882.40	743.35	29.4	29.99	34.99	1	5000	0	LISWYDNEFGYSNR
887.41	1111.47	29.4	29.99	34.99	1	5000	0	LISWYDNEFGYSNR
887.41	996.44	29.4	29.99	34.99	1	5000	0	LISWYDNEFGYSNR
887.41	753.36	29.4	29.99	34.99	1	5000	0	LISWYDNEFGYSNR
847.95	1201.57	28.3	25.51	30.51	1	5000	0	LIAPVAEEEATVPNNK
847.95	571.32	28.3	25.51	30.51	1	5000	0	LIAPVAEEEATVPNNK
847.95	472.25	28.3	25.51	30.51	1	5000	0	LIAPVAEEEATVPNNK
851.96	1209.58	28.3	25.51	30.51	1	5000	0	LIAPVAEEEATVPNNK
851.96	579.33	28.3	25.51	30.51	1	5000	0	LIAPVAEEEATVPNNK
851.96	480.27	28.3	25.51	30.51	1	5000	0	LIAPVAEEEATVPNNK
815.43	1001.55	27.4	35.22	40.22	1	5000	0	SLADELALVDVLEDK
815.43	930.51	27.4	35.22	40.22	1	5000	0	SLADELALVDVLEDK
815.43	817.43	27.4	35.22	40.22	1	5000	0	SLADELALVDVLEDK
819.44	1009.57	27.4	35.22	40.22	1	5000	0	SLADELALVDVLEDK
819.44	938.53	27.4	35.22	40.22	1	5000	0	SLADELALVDVLEDK
819.44	825.44	27.4	35.22	40.22	1	5000	0	SLADELALVDVLEDK
457.30	701.43	16.6	23.78	28.78	1	5000	0	IVVVTAGVR
457.30	602.36	16.6	23.78	28.78	1	5000	0	IVVVTAGVR
457.30	503.29	16.6	23.78	28.78	1	5000	0	IVVVTAGVR
462.30	711.44	16.6	23.78	28.78	1	5000	0	IVVVTAGVR
462.30	612.37	16.6	23.78	28.78	1	5000	0	IVVVTAGVR
462.30	513.30	16.6	23.78	28.78	1	5000	0	IVVVTAGVR
596.32	831.49	24.9	27.5	32.5	1	5000	0	HLEINPDHSIIETLR
596.32	744.46	24.9	27.5	32.5	1	5000	0	HLEINPDHSIIETLR
596.32	631.38	24.9	27.5	32.5	1	5000	0	HLEINPDHSIIETLR
599.66	841.50	24.9	27.5	32.5	1	5000	0	HLEINPDHSIIETLR
599.66	754.47	24.9	27.5	32.5	1	5000	0	HLEINPDHSIIETLR
599.66	641.39	24.9	27.5	32.5	1	5000	0	HLEINPDHSIIETLR
618.30	992.52	21.5	22.73	27.73	1	5000	0	DQVANSAFVER
618.30	893.45	21.5	22.73	27.73	1	5000	0	DQVANSAFVER
623.31	1002.52	21.5	22.73	27.73	1	5000	0	DQVANSAFVER
623.31	903.46	21.5	22.73	27.73	1	5000	0	DQVANSAFVER
904.40	1344.58	30	23.92	28.92	1	5000	0	NSYVAGQYDDAASYQR
904.40	1088.46	30	23.92	28.92	1	5000	0	NSYVAGQYDDAASYQR
904.40	925.40	30	23.92	28.92	1	5000	0	NSYVAGQYDDAASYQR
909.40	1354.59	30	23.92	28.92	1	5000	0	NSYVAGQYDDAASYQR
909.40	1098.47	30	23.92	28.92	1	5000	0	NSYVAGQYDDAASYQR
909.40	935.41	30	23.92	28.92	1	5000	0	NSYVAGQYDDAASYQR
637.31	996.51	22	31.07	36.07	1	5000	0	GGYFDEFGIIR
637.31	849.45	22	31.07	36.07	1	5000	0	GGYFDEFGIIR
637.31	605.38	22	31.07	36.07	1	5000	0	GGYFDEFGIIR

Supplementary table 1 contd.

642.32	1006.52	22	31.07	36.07	1	5000	0	GGYFDEFGIIR
642.32	859.45	22	31.07	36.07	1	5000	0	GGYFDEFGIIR
642.32	615.39	22	31.07	36.07	1	5000	0	GGYFDEFGIIR
891.42	1144.58	29.6	28.29	33.29	1	5000	0	DYPDFSPSVDAEAIQK
891.42	1057.55	29.6	28.29	33.29	1	5000	0	DYPDFSPSVDAEAIQK
895.42	1152.60	29.6	28.29	33.29	1	5000	0	DYPDFSPSVDAEAIQK
895.42	1065.57	29.6	28.29	33.29	1	5000	0	DYPDFSPSVDAEAIQK
721.36	1138.60	24.5	27.96	32.96	1	5000	0	SDTSGDYEITLLK
721.36	1051.57	24.5	27.96	32.96	1	5000	0	SDTSGDYEITLLK
725.36	1146.61	24.5	27.96	32.96	1	5000	0	SDTSGDYEITLLK
725.36	1059.58	24.5	27.96	32.96	1	5000	0	SDTSGDYEITLLK
472.23	729.32	17.1	24.29	29.29	1	5000	0	LTFDEYR
472.23	582.25	17.1	24.29	29.29	1	5000	0	LTFDEYR
477.23	739.33	17.1	24.29	29.29	1	5000	0	LTFDEYR
477.23	592.26	17.1	24.29	29.29	1	5000	0	LTFDEYR
556.80	928.51	19.6	20.5	25.5	1	5000	0	SPPGQVTEAVK
556.80	831.46	19.6	20.5	25.5	1	5000	0	SPPGQVTEAVK
556.80	646.38	19.6	20.5	25.5	1	5000	0	SPPGQVTEAVK
560.81	936.52	19.6	20.5	25.5	1	5000	0	SPPGQVTEAVK
560.81	839.47	19.6	20.5	25.5	1	5000	0	SPPGQVTEAVK
560.81	654.39	19.6	20.5	25.5	1	5000	0	SPPGQVTEAVK
558.32	887.53	19.7	31.13	36.13	1	5000	0	MPILGLGTWK
558.32	774.45	19.7	31.13	36.13	1	5000	0	MPILGLGTWK
558.32	661.37	19.7	31.13	36.13	1	5000	0	MPILGLGTWK
562.32	895.55	19.7	31.13	36.13	1	5000	0	MPILGLGTWK
562.32	782.47	19.7	31.13	36.13	1	5000	0	MPILGLGTWK
562.32	669.38	19.7	31.13	36.13	1	5000	0	MPILGLGTWK
446.75	722.38	16.3	23.89	28.89	1	5000	0	VAIDVGYR
446.75	609.30	16.3	23.89	28.89	1	5000	0	VAIDVGYR
446.75	494.27	16.3	23.89	28.89	1	5000	0	VAIDVGYR
451.75	732.39	16.3	23.89	28.89	1	5000	0	VAIDVGYR
451.75	619.31	16.3	23.89	28.89	1	5000	0	VAIDVGYR
451.75	504.28	16.3	23.89	28.89	1	5000	0	VAIDVGYR
824.87	1182.52	27.7	30.13	35.13	1	5000	0	FGYVDFESAEDLEK
824.87	1067.49	27.7	30.13	35.13	1	5000	0	FGYVDFESAEDLEK
824.87	791.38	27.7	30.13	35.13	1	5000	0	FGYVDFESAEDLEK
828.88	1190.53	27.7	30.13	35.13	1	5000	0	FGYVDFESAEDLEK
828.88	1075.50	27.7	30.13	35.13	1	5000	0	FGYVDFESAEDLEK
828.88	799.39	27.7	30.13	35.13	1	5000	0	FGYVDFESAEDLEK
589.79	950.46	20.6	25.28	30.28	1	5000	0	EVFEDAAEIR
589.79	803.39	20.6	25.28	30.28	1	5000	0	EVFEDAAEIR
589.79	674.35	20.6	25.28	30.28	1	5000	0	EVFEDAAEIR
594.79	960.47	20.6	25.28	30.28	1	5000	0	EVFEDAAEIR
594.79	813.40	20.6	25.28	30.28	1	5000	0	EVFEDAAEIR
594.79	684.36	20.6	25.28	30.28	1	5000	0	EVFEDAAEIR
781.34	1153.50	26.3	29.72	34.72	1	5000	0	GFGFVDFNSEEDAK
781.34	1054.43	26.3	29.72	34.72	1	5000	0	GFGFVDFNSEEDAK
781.34	939.41	26.3	29.72	34.72	1	5000	0	GFGFVDFNSEEDAK
785.35	1161.51	26.3	29.72	34.72	1	5000	0	GFGFVDFNSEEDAK
785.35	1062.45	26.3	29.72	34.72	1	5000	0	GFGFVDFNSEEDAK
785.35	947.42	26.3	29.72	34.72	1	5000	0	GFGFVDFNSEEDAK
696.87	933.50	23.8	26.65	31.65	1	5000	0	STGTISVISSGLDR
696.87	747.40	23.8	26.65	31.65	1	5000	0	STGTISVISSGLDR
696.87	634.32	23.8	26.65	31.65	1	5000	0	STGTISVISSGLDR
701.87	943.51	23.8	26.65	31.65	1	5000	0	STGTISVISSGLDR
701.87	757.41	23.8	26.65	31.65	1	5000	0	STGTISVISSGLDR
701.87	644.32	23.8	26.65	31.65	1	5000	0	STGTISVISSGLDR
828.42	1013.48	27.8	28.53	33.53	1	5000	0	LTVTDLDAPNSPAWR
828.42	898.45	27.8	28.53	33.53	1	5000	0	LTVTDLDAPNSPAWR
828.42	827.42	27.8	28.53	33.53	1	5000	0	LTVTDLDAPNSPAWR
833.43	1023.49	27.8	28.53	33.53	1	5000	0	LTVTDLDAPNSPAWR
833.43	908.46	27.8	28.53	33.53	1	5000	0	LTVTDLDAPNSPAWR
833.43	837.42	27.8	28.53	33.53	1	5000	0	LTVTDLDAPNSPAWR
893.91	1329.63	29.7	23.57	28.57	1	5000	0	EAAGEGPALYEDPPDQK
893.91	991.44	29.7	23.57	28.57	1	5000	0	EAAGEGPALYEDPPDQK
893.91	584.30	29.7	23.57	28.57	1	5000	0	EAAGEGPALYEDPPDQK
897.92	1337.65	29.7	23.57	28.57	1	5000	0	EAAGEGPALYEDPPDQK
897.92	999.45	29.7	23.57	28.57	1	5000	0	EAAGEGPALYEDPPDQK
897.92	592.32	29.7	23.57	28.57	1	5000	0	EAAGEGPALYEDPPDQK
924.49	1103.58	30.6	35.51	40.51	1	5000	0	QGFGELLQAVPLADSFR
924.49	975.53	30.6	35.51	40.51	1	5000	0	QGFGELLQAVPLADSFR
924.49	805.42	30.6	35.51	40.51	1	5000	0	QGFGELLQAVPLADSFR

Supplementary table 1 contd.

929.49	1113.59	30.6	35.51	40.51	1	5000	0	QGFGELLQAVPLADSFR
929.49	985.53	30.6	35.51	40.51	1	5000	0	QGFGELLQAVPLADSFR
929.49	815.43	30.6	35.51	40.51	1	5000	0	QGFGELLQAVPLADSFR
1046.54	790.38	34.3	34.88	39.88	1	5000	0	LDYFLLSHSLLPALC[+57.0]DSK
1046.54	622.29	34.3	34.88	39.88	1	5000	0	LDYFLLSHSLLPALC[+57.0]DSK
1046.54	392.18	34.3	34.88	39.88	1	5000	0	LDYFLLSHSLLPALC[+57.0]DSK
1050.55	798.39	34.3	34.88	39.88	1	5000	0	LDYFLLSHSLLPALC[+57.0]DSK
1050.55	630.30	34.3	34.88	39.88	1	5000	0	LDYFLLSHSLLPALC[+57.0]DSK
1050.55	392.18	34.3	34.88	39.88	1	5000	0	LDYFLLSHSLLPALC[+57.0]DSK
510.25	906.42	18.2	23.11	28.11	1	5000	0	ISSPTETER
510.25	819.38	18.2	23.11	28.11	1	5000	0	ISSPTETER
510.25	732.35	18.2	23.11	28.11	1	5000	0	ISSPTETER
515.26	916.42	18.2	23.11	28.11	1	5000	0	ISSPTETER
515.26	829.39	18.2	23.11	28.11	1	5000	0	ISSPTETER
515.26	742.36	18.2	23.11	28.11	1	5000	0	ISSPTETER
530.75	725.38	18.8	22.9	27.9	1	5000	0	DGYNYTLSK
530.75	611.34	18.8	22.9	27.9	1	5000	0	DGYNYTLSK
530.75	448.28	18.8	22.9	27.9	1	5000	0	DGYNYTLSK
534.76	733.40	18.8	22.9	27.9	1	5000	0	DGYNYTLSK
534.76	619.35	18.8	22.9	27.9	1	5000	0	DGYNYTLSK
534.76	456.29	18.8	22.9	27.9	1	5000	0	DGYNYTLSK
386.20	559.32	14.5	21.14	26.14	1	5000	0	DPGVLDR
386.20	502.30	14.5	21.14	26.14	1	5000	0	DPGVLDR
386.20	403.23	14.5	21.14	26.14	1	5000	0	DPGVLDR
391.21	569.33	14.5	21.14	26.14	1	5000	0	DPGVLDR
391.21	512.31	14.5	21.14	26.14	1	5000	0	DPGVLDR
391.21	413.24	14.5	21.14	26.14	1	5000	0	DPGVLDR
551.32	746.42	19.4	28.44	33.44	1	5000	0	IIELPFQNK
551.32	633.34	19.4	28.44	33.44	1	5000	0	IIELPFQNK
555.33	754.43	19.4	28.44	33.44	1	5000	0	IIELPFQNK
555.33	641.35	19.4	28.44	33.44	1	5000	0	IIELPFQNK
473.25	703.37	17.1	22.02	27.02	1	5000	0	ELVLDNSR
473.25	604.30	17.1	22.02	27.02	1	5000	0	ELVLDNSR
473.25	491.22	17.1	22.02	27.02	1	5000	0	ELVLDNSR
478.26	713.38	17.1	22.02	27.02	1	5000	0	ELVLDNSR
478.26	614.31	17.1	22.02	27.02	1	5000	0	ELVLDNSR
478.26	501.23	17.1	22.02	27.02	1	5000	0	ELVLDNSR
489.92	745.42	20.9	24.6	29.6	1	5000	0	C[+57.0]PNLTHLNLSGNK
489.92	632.34	20.9	24.6	29.6	1	5000	0	C[+57.0]PNLTHLNLSGNK
489.92	405.21	20.9	24.6	29.6	1	5000	0	C[+57.0]PNLTHLNLSGNK
492.59	753.43	20.9	24.6	29.6	1	5000	0	C[+57.0]PNLTHLNLSGNK
492.59	640.35	20.9	24.6	29.6	1	5000	0	C[+57.0]PNLTHLNLSGNK
492.59	413.22	20.9	24.6	29.6	1	5000	0	C[+57.0]PNLTHLNLSGNK
982.01	1295.66	32.4	26.97	31.97	1	5000	0	VELPGTAVPSVPEDAAPASR
982.01	1196.59	32.4	26.97	31.97	1	5000	0	VELPGTAVPSVPEDAAPASR
982.01	913.44	32.4	26.97	31.97	1	5000	0	VELPGTAVPSVPEDAAPASR
987.01	1305.67	32.4	26.97	31.97	1	5000	0	VELPGTAVPSVPEDAAPASR
987.01	1206.60	32.4	26.97	31.97	1	5000	0	VELPGTAVPSVPEDAAPASR
987.01	923.45	32.4	26.97	31.97	1	5000	0	VELPGTAVPSVPEDAAPASR
687.84	993.51	23.5	25.18	30.18	1	5000	0	AFYAPVHADDLR
687.84	922.47	23.5	25.18	30.18	1	5000	0	AFYAPVHADDLR
687.84	726.35	23.5	25.18	30.18	1	5000	0	AFYAPVHADDLR
692.85	1003.52	23.5	25.18	30.18	1	5000	0	AFYAPVHADDLR
692.85	932.48	23.5	25.18	30.18	1	5000	0	AFYAPVHADDLR
692.85	736.36	23.5	25.18	30.18	1	5000	0	AFYAPVHADDLR
458.90	922.47	19.7	25.18	30.18	1	5000	0	AFYAPVHADDLR
458.90	726.35	19.7	25.18	30.18	1	5000	0	AFYAPVHADDLR
458.90	589.29	19.7	25.18	30.18	1	5000	0	AFYAPVHADDLR
462.23	932.48	19.7	25.18	30.18	1	5000	0	AFYAPVHADDLR
462.23	736.36	19.7	25.18	30.18	1	5000	0	AFYAPVHADDLR
462.23	599.30	19.7	25.18	30.18	1	5000	0	AFYAPVHADDLR
747.37	804.45	30.7	28.6	33.6	1	5000	0	EGLDISHLQGQEELLSSQEK
747.37	691.36	30.7	28.6	33.6	1	5000	0	EGLDISHLQGQEELLSSQEK
747.37	578.28	30.7	28.6	33.6	1	5000	0	EGLDISHLQGQEELLSSQEK
750.04	812.46	30.7	28.6	33.6	1	5000	0	EGLDISHLQGQEELLSSQEK
750.04	699.38	30.7	28.6	33.6	1	5000	0	EGLDISHLQGQEELLSSQEK
750.04	586.29	30.7	28.6	33.6	1	5000	0	EGLDISHLQGQEELLSSQEK
619.32	1124.55	21.5	26.04	31.04	1	5000	0	ITDNELELYK
619.32	1023.50	21.5	26.04	31.04	1	5000	0	ITDNELELYK
619.32	310.18	21.5	26.04	31.04	1	5000	0	ITDNELELYK
623.33	1132.56	21.5	26.04	31.04	1	5000	0	ITDNELELYK
623.33	1031.51	21.5	26.04	31.04	1	5000	0	ITDNELELYK

Supplementary table 1 contd.

623.33	318.19	21.5	26.04	31.04	1	5000	0	ITDNELELYK
632.80	1003.48	21.9	27.78	32.78	1	5000	0	FNVWDVGGQDK
632.80	904.42	21.9	27.78	32.78	1	5000	0	FNVWDVGGQDK
632.80	718.34	21.9	27.78	32.78	1	5000	0	FNVWDVGGQDK
636.81	1011.50	21.9	27.78	32.78	1	5000	0	FNVWDVGGQDK
636.81	912.43	21.9	27.78	32.78	1	5000	0	FNVWDVGGQDK
636.81	726.35	21.9	27.78	32.78	1	5000	0	FNVWDVGGQDK
628.34	1005.57	21.8	28.96	33.96	1	5000	0	SYPGLTSYLVR
628.34	908.52	21.8	28.96	33.96	1	5000	0	SYPGLTSYLVR
633.34	1015.58	21.8	28.96	33.96	1	5000	0	SYPGLTSYLVR
633.34	918.53	21.8	28.96	33.96	1	5000	0	SYPGLTSYLVR
975.52	1333.75	32.2	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
975.52	1220.66	32.2	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
975.52	994.57	32.2	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
980.52	1343.76	32.2	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
980.52	1230.67	32.2	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
980.52	1004.58	32.2	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
650.68	994.57	27	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
650.68	897.52	27	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
650.68	840.49	27	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
654.02	1004.58	27	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
654.02	907.52	27	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
654.02	850.50	27	30.44	35.44	1	5000	0	LENYPIPEPGPNEVLLR
791.92	1226.64	26.7	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
791.92	1127.57	26.7	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
791.92	1014.49	26.7	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
795.92	1234.65	26.7	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
795.92	1135.58	26.7	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
795.92	1022.50	26.7	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
528.28	820.40	22.4	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
528.28	691.36	22.4	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
528.28	620.32	22.4	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
530.95	828.42	22.4	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
530.95	699.38	22.4	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
530.95	628.34	22.4	20.5	25.5	1	5000	0	KPMVLGHEASGTVEK
879.78	1351.72	35.7	33.18	38.18	1	5000	0	LPDNVTFEEGALIEPLSVGIHAC[+57.0]R
879.78	1238.63	35.7	33.18	38.18	1	5000	0	LPDNVTFEEGALIEPLSVGIHAC[+57.0]R
879.78	1109.59	35.7	33.18	38.18	1	5000	0	LPDNVTFEEGALIEPLSVGIHAC[+57.0]R
883.12	1361.72	35.7	33.18	38.18	1	5000	0	LPDNVTFEEGALIEPLSVGIHAC[+57.0]R
883.12	1248.64	35.7	33.18	38.18	1	5000	0	LPDNVTFEEGALIEPLSVGIHAC[+57.0]R
883.12	1119.60	35.7	33.18	38.18	1	5000	0	LPDNVTFEEGALIEPLSVGIHAC[+57.0]R
795.42	961.53	26.8	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
795.42	862.46	26.8	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
795.42	763.39	26.8	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
800.42	971.54	26.8	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
800.42	872.47	26.8	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
800.42	773.40	26.8	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
530.62	434.24	22.4	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
530.62	728.38	22.4	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
530.62	1043.52	22.4	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
533.95	444.24	22.4	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
533.95	728.38	22.4	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
533.95	1043.52	22.4	29.11	34.11	1	5000	0	AMGAAQVVVTDLSATR
643.37	1043.61	22.2	30.34	35.34	1	5000	0	EIGADLVLQISK
643.37	687.44	22.2	30.34	35.34	1	5000	0	EIGADLVLQISK
643.37	588.37	22.2	30.34	35.34	1	5000	0	EIGADLVLQISK
647.38	1051.62	22.2	30.34	35.34	1	5000	0	EIGADLVLQISK
647.38	695.45	22.2	30.34	35.34	1	5000	0	EIGADLVLQISK
647.38	596.39	22.2	30.34	35.34	1	5000	0	EIGADLVLQISK
487.26	860.42	17.5	17.5	22.5	1	5000	0	LGGNEQVTR
487.26	803.40	17.5	17.5	22.5	1	5000	0	LGGNEQVTR
487.26	503.29	17.5	17.5	22.5	1	5000	0	LGGNEQVTR
644.82	1016.53	22.2	19.75	24.75	1	5000	0	AGGSSEPVTGLADK
644.82	800.45	22.2	19.75	24.75	1	5000	0	AGGSSEPVTGLADK
644.82	604.33	22.2	19.75	24.75	1	5000	0	AGGSSEPVTGLADK
683.83	819.39	23.4	22.24	27.24	1	5000	0	VEATFGVDESANK
683.83	966.45	23.4	22.24	27.24	1	5000	0	VEATFGVDESANK
683.83	663.29	23.4	22.24	27.24	1	5000	0	VEATFGVDESANK
547.30	817.44	19.3	23.89	28.89	1	5000	0	YILAGVENSK
547.30	704.36	19.3	23.89	28.89	1	5000	0	YILAGVENSK
547.30	633.32	19.3	23.89	28.89	1	5000	0	YILAGVENSK
669.84	1041.50	23	24.62	29.62	1	5000	0	TPVISGGPYYER

Supplementary table 1 contd.

669.84	928.42	23	24.62	29.62	1	5000	0	TPVISGGPYYER
669.84	841.38	23	24.62	29.62	1	5000	0	TPVISGGPYYER
683.85	1069.53	23.4	25.32	30.32	1	5000	0	TPVITGAPYYER
683.85	956.45	23.4	25.32	30.32	1	5000	0	TPVITGAPYYER
683.85	855.40	23.4	25.32	30.32	1	5000	0	TPVITGAPYYER
699.34	926.47	23.9	27.02	32.02	1	5000	0	GDLDAASYYAPVR
699.34	855.44	23.9	27.02	32.02	1	5000	0	GDLDAASYYAPVR
699.34	605.34	23.9	27.02	32.02	1	5000	0	GDLDAASYYAPVR
726.84	1066.48	24.7	27.94	32.94	1	5000	0	DAVTPADFSEWSK
726.84	584.27	24.7	27.94	32.94	1	5000	0	DAVTPADFSEWSK
726.84	533.75	24.7	27.94	32.94	1	5000	0	DAVTPADFSEWSK
622.85	826.48	21.6	30.57	35.57	1	5000	0	TGFIIDPGGVIR
622.85	713.39	21.6	30.57	35.57	1	5000	0	TGFIIDPGGVIR
622.85	598.37	21.6	30.57	35.57	1	5000	0	TGFIIDPGGVIR
636.87	854.51	22	32.8	37.8	1	5000	0	GTFIIDPAAIVR
636.87	741.43	22	32.8	37.8	1	5000	0	GTFIIDPAAIVR
636.87	626.40	22	32.8	37.8	1	5000	0	GTFIIDPAAIVR
776.93	1051.56	26.2	34.1	39.1	1	5000	0	FLLQFGAQGSPLFK
776.93	904.49	26.2	34.1	39.1	1	5000	0	FLLQFGAQGSPLFK
776.93	504.32	26.2	34.1	39.1	1	5000	0	FLLQFGAQGSPLFK

Supplementary table 2. List of transitions selected for scheduled SRM on the TSQ Vantage (Method 2)

Q1	Q3	CE	RT-2.5	RT+2.5	Polarity	Trigger	Reference	Peptide Sequence
487.26	860.42	17.5	17.22	22.22	1	5000	0	LGGNEQVTR
487.26	803.40	17.5	17.22	22.22	1	5000	0	LGGNEQVTR
487.26	503.29	17.5	17.22	22.22	1	5000	0	LGGNEQVTR
644.82	1016.53	22.2	19.69	24.69	1	5000	0	AGGSSEPVTGLADK
644.82	800.45	22.2	19.69	24.69	1	5000	0	AGGSSEPVTGLADK
644.82	604.33	22.2	19.69	24.69	1	5000	0	AGGSSEPVTGLADK
683.83	819.39	23.4	22.08	27.08	1	5000	0	VEATFGVDESANK
683.83	966.45	23.4	22.08	27.08	1	5000	0	VEATFGVDESANK
683.83	663.29	23.4	22.08	27.08	1	5000	0	VEATFGVDESANK
547.30	817.44	19.3	23.66	28.66	1	5000	0	YILAGVENSK
547.30	704.36	19.3	23.66	28.66	1	5000	0	YILAGVENSK
547.30	633.32	19.3	23.66	28.66	1	5000	0	YILAGVENSK
669.84	1041.50	23	24.3	29.3	1	5000	0	TPVISGGPYYER
669.84	928.42	23	24.3	29.3	1	5000	0	TPVISGGPYYER
669.84	841.38	23	24.3	29.3	1	5000	0	TPVISGGPYYER
683.85	1069.53	23.4	25.08	30.08	1	5000	0	TPVITGAPYYER
683.85	956.45	23.4	25.08	30.08	1	5000	0	TPVITGAPYYER
683.85	855.40	23.4	25.08	30.08	1	5000	0	TPVITGAPYYER
699.34	926.47	23.9	26.34	31.34	1	5000	0	GDLDAASYYAPVR
699.34	855.44	23.9	26.34	31.34	1	5000	0	GDLDAASYYAPVR
699.34	605.34	23.9	26.34	31.34	1	5000	0	GDLDAASYYAPVR
726.84	1066.48	24.7	27.73	32.73	1	5000	0	DAVTPADFSEWSK
726.84	584.27	24.7	27.73	32.73	1	5000	0	DAVTPADFSEWSK
726.84	533.75	24.7	27.73	32.73	1	5000	0	DAVTPADFSEWSK
622.85	826.48	21.6	30.59	35.59	1	5000	0	TGFIIDPGGVIR
622.85	713.39	21.6	30.59	35.59	1	5000	0	TGFIIDPGGVIR
622.85	598.37	21.6	30.59	35.59	1	5000	0	TGFIIDPGGVIR
636.87	854.51	22	32.9	37.9	1	5000	0	GTFIIDPAAIVR
636.87	741.43	22	32.9	37.9	1	5000	0	GTFIIDPAAIVR
636.87	626.40	22	32.9	37.9	1	5000	0	GTFIIDPAAIVR
776.93	1051.56	26.2	34	39	1	5000	0	FLLQFGAQGSPLFK
776.93	904.49	26.2	34	39	1	5000	0	FLLQFGAQGSPLFK
776.93	504.32	26.2	34	39	1	5000	0	FLLQFGAQGSPLFK
604.81	980.49	21	26.62	31.62	1	5000	0	VEVTEFEDIK
604.81	881.43	21	26.62	31.62	1	5000	0	VEVTEFEDIK
604.81	651.33	21	26.62	31.62	1	5000	0	VEVTEFEDIK
608.81	988.51	21	26.62	31.62	1	5000	0	VEVTEFEDIK
608.81	889.44	21	26.62	31.62	1	5000	0	VEVTEFEDIK
608.81	659.35	21	26.62	31.62	1	5000	0	VEVTEFEDIK
920.91	1302.56	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
920.91	1155.50	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
920.91	797.38	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
920.91	537.27	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
924.91	1310.58	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
924.91	1163.51	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
924.91	805.40	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
924.91	545.28	30.5	32.32	37.32	1	5000	0	IDFYFDENPYFENK
723.83	1033.48	24.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
723.83	920.40	24.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
723.83	418.23	24.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
727.84	1041.49	24.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
727.84	928.41	24.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
727.84	426.24	24.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
482.89	590.28	20.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
482.89	418.23	20.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
482.89	321.18	20.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
485.56	598.29	20.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
485.56	426.24	20.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
485.56	329.19	20.6	19.12	24.12	1	5000	0	EFHLNESGDPSSK
637.83	1078.53	22	26.11	31.11	1	5000	0	VPVYETPAGWR
637.83	979.46	22	26.11	31.11	1	5000	0	VPVYETPAGWR
637.83	816.40	22	26.11	31.11	1	5000	0	VPVYETPAGWR
642.83	1088.54	22	26.11	31.11	1	5000	0	VPVYETPAGWR
642.83	989.47	22	26.11	31.11	1	5000	0	VPVYETPAGWR

Supplementary table 2 contd.

642.83	826.41	22	26.11	31.11	1	5000	0	VPVYETPAGWR
527.27	911.46	18.7	26.59	31.59	1	5000	0	AAPFSLEYR
527.27	814.41	18.7	26.59	31.59	1	5000	0	AAPFSLEYR
527.27	667.34	18.7	26.59	31.59	1	5000	0	AAPFSLEYR
532.28	921.47	18.7	26.59	31.59	1	5000	0	AAPFSLEYR
532.28	824.42	18.7	26.59	31.59	1	5000	0	AAPFSLEYR
532.28	677.35	18.7	26.59	31.59	1	5000	0	AAPFSLEYR
932.47	706.38	30.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
932.47	496.24	30.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
932.47	1158.56	30.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
936.47	714.39	30.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
936.47	504.25	30.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
936.47	1158.56	30.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
621.98	934.49	25.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
621.98	706.38	25.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
621.98	496.24	25.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
624.65	942.50	25.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
624.65	714.39	25.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
624.65	504.25	25.9	29.75	34.75	1	5000	0	GQYISPFHDIPIYADK
664.34	829.46	22.8	29.63	34.63	1	5000	0	DVFHMVVEVPR
664.34	698.42	22.8	29.63	34.63	1	5000	0	DVFHMVVEVPR
664.34	599.35	22.8	29.63	34.63	1	5000	0	DVFHMVVEVPR
669.35	839.47	22.8	29.63	34.63	1	5000	0	DVFHMVVEVPR
669.35	708.43	22.8	29.63	34.63	1	5000	0	DVFHMVVEVPR
669.35	609.36	22.8	29.63	34.63	1	5000	0	DVFHMVVEVPR
443.23	829.46	19.1	29.63	34.63	1	5000	0	DVFHMVVEVPR
443.23	599.35	19.1	29.63	34.63	1	5000	0	DVFHMVVEVPR
443.23	500.28	19.1	29.63	34.63	1	5000	0	DVFHMVVEVPR
446.57	839.47	19.1	29.63	34.63	1	5000	0	DVFHMVVEVPR
446.57	609.36	19.1	29.63	34.63	1	5000	0	DVFHMVVEVPR
446.57	510.29	19.1	29.63	34.63	1	5000	0	DVFHMVVEVPR
650.65	324.15	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
650.65	897.42	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
650.65	782.39	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
650.65	574.29	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
653.98	324.15	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
653.98	907.43	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
653.98	792.40	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
653.98	584.30	27	25.1	30.1	1	5000	0	HWILPQDYDHAQAEAR
485.26	806.44	17.5	24.74	29.74	1	5000	0	YIFDNVAK
485.26	693.36	17.5	24.74	29.74	1	5000	0	YIFDNVAK
485.26	546.29	17.5	24.74	29.74	1	5000	0	YIFDNVAK
485.26	431.26	17.5	24.74	29.74	1	5000	0	YIFDNVAK
489.26	814.45	17.5	24.74	29.74	1	5000	0	YIFDNVAK
489.26	701.37	17.5	24.74	29.74	1	5000	0	YIFDNVAK
489.26	554.30	17.5	24.74	29.74	1	5000	0	YIFDNVAK
489.26	439.28	17.5	24.74	29.74	1	5000	0	YIFDNVAK
850.92	1151.58	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
850.92	1052.52	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
850.92	905.45	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
850.92	834.41	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
855.92	1161.59	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
855.92	1062.52	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
855.92	915.46	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
855.92	844.42	28.4	29.83	34.83	1	5000	0	EGVYTVFAPTNEAFR
658.40	1101.67	22.7	34.29	39.29	1	5000	0	LTLLAPLNSVFK
658.40	988.58	22.7	34.29	39.29	1	5000	0	LTLLAPLNSVFK
658.40	875.50	22.7	34.29	39.29	1	5000	0	LTLLAPLNSVFK
662.41	1109.68	22.7	34.29	39.29	1	5000	0	LTLLAPLNSVFK
662.41	996.60	22.7	34.29	39.29	1	5000	0	LTLLAPLNSVFK
662.41	883.51	22.7	34.29	39.29	1	5000	0	LTLLAPLNSVFK
608.38	861.52	21.2	34.25	39.25	1	5000	0	VALAGLLGFGLGK
608.38	691.41	21.2	34.25	39.25	1	5000	0	VALAGLLGFGLGK
608.38	578.33	21.2	34.25	39.25	1	5000	0	VALAGLLGFGLGK
612.38	869.53	21.2	34.25	39.25	1	5000	0	VALAGLLGFGLGK
612.38	699.43	21.2	34.25	39.25	1	5000	0	VALAGLLGFGLGK
612.38	586.34	21.2	34.25	39.25	1	5000	0	VALAGLLGFGLGK
332.50	428.25	14.9	19.45	24.45	1	5000	0	DAHFPPPSK

Supplementary table 2 contd.

332.50	331.20	14.9	19.45	24.45	1	5000	0	DAHFPPPSK
332.50	471.20	14.9	19.45	24.45	1	5000	0	DAHFPPPSK
332.50	568.25	14.9	19.45	24.45	1	5000	0	DAHFPPPSK
335.17	436.26	14.9	19.45	24.45	1	5000	0	DAHFPPPSK
335.17	339.21	14.9	19.45	24.45	1	5000	0	DAHFPPPSK
335.17	471.20	14.9	19.45	24.45	1	5000	0	DAHFPPPSK
335.17	568.25	14.9	19.45	24.45	1	5000	0	DAHFPPPSK
731.38	942.50	24.8	29.83	34.83	1	5000	0	GLYWVAPLNTDGR
731.38	843.43	24.8	29.83	34.83	1	5000	0	GLYWVAPLNTDGR
736.38	952.51	24.8	29.83	34.83	1	5000	0	GLYWVAPLNTDGR
736.38	853.44	24.8	29.83	34.83	1	5000	0	GLYWVAPLNTDGR
638.81	822.43	22.1	29.83	34.83	1	5000	0	GFSYLYGAWGR
638.81	709.34	22.1	29.83	34.83	1	5000	0	GFSYLYGAWGR
638.81	546.28	22.1	29.83	34.83	1	5000	0	GFSYLYGAWGR
643.81	832.43	22.1	29.83	34.83	1	5000	0	GFSYLYGAWGR
643.81	719.35	22.1	29.83	34.83	1	5000	0	GFSYLYGAWGR
643.81	556.29	22.1	29.83	34.83	1	5000	0	GFSYLYGAWGR
628.36	1041.58	21.8	28.48	33.48	1	5000	0	TLPIGQNFPIR
628.36	944.53	21.8	28.48	33.48	1	5000	0	TLPIGQNFPIR
628.36	831.45	21.8	28.48	33.48	1	5000	0	TLPIGQNFPIR
633.37	1051.59	21.8	28.48	33.48	1	5000	0	TLPIGQNFPIR
633.37	954.54	21.8	28.48	33.48	1	5000	0	TLPIGQNFPIR
633.37	841.46	21.8	28.48	33.48	1	5000	0	TLPIGQNFPIR
607.32	784.49	25.4	24.53	29.53	1	5000	0	STHYQQYQPVVTLQK
607.32	588.37	25.4	24.53	29.53	1	5000	0	STHYQQYQPVVTLQK
607.32	489.30	25.4	24.53	29.53	1	5000	0	STHYQQYQPVVTLQK
609.99	792.51	25.4	24.53	29.53	1	5000	0	STHYQQYQPVVTLQK
609.99	596.39	25.4	24.53	29.53	1	5000	0	STHYQQYQPVVTLQK
609.99	497.32	25.4	24.53	29.53	1	5000	0	STHYQQYQPVVTLQK
619.90	751.51	21.5	30.79	35.79	1	5000	0	IFQVVPIPVVK
619.90	652.44	21.5	30.79	35.79	1	5000	0	IFQVVPIPVVK
619.90	442.30	21.5	30.79	35.79	1	5000	0	IFQVVPIPVVK
623.90	759.52	21.5	30.79	35.79	1	5000	0	IFQVVPIPVVK
623.90	660.45	21.5	30.79	35.79	1	5000	0	IFQVVPIPVVK
623.90	450.32	21.5	30.79	35.79	1	5000	0	IFQVVPIPVVK
495.75	748.37	17.8	21.94	26.94	1	5000	0	ELVLDNC[+57.0]K
495.75	649.30	17.8	21.94	26.94	1	5000	0	ELVLDNC[+57.0]K
495.75	536.21	17.8	21.94	26.94	1	5000	0	ELVLDNC[+57.0]K
499.76	756.38	17.8	21.94	26.94	1	5000	0	ELVLDNC[+57.0]K
499.76	657.31	17.8	21.94	26.94	1	5000	0	ELVLDNC[+57.0]K
499.76	544.23	17.8	21.94	26.94	1	5000	0	ELVLDNC[+57.0]K
474.27	745.42	20.3	26.92	31.92	1	5000	0	LPNLTHLNLSGNK
474.27	632.34	20.3	26.92	31.92	1	5000	0	LPNLTHLNLSGNK
474.27	405.21	20.3	26.92	31.92	1	5000	0	LPNLTHLNLSGNK
476.94	753.43	20.3	26.92	31.92	1	5000	0	LPNLTHLNLSGNK
476.94	640.35	20.3	26.92	31.92	1	5000	0	LPNLTHLNLSGNK
476.94	413.22	20.3	26.92	31.92	1	5000	0	LPNLTHLNLSGNK
728.39	989.52	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
728.39	861.46	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
728.39	707.38	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
728.39	553.31	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
733.39	999.52	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
733.39	871.47	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
733.39	717.39	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
733.39	563.32	24.8	19.46	24.46	1	5000	0	LGAPQQPGPGPPPSR
747.88	1002.56	25.3	32.04	37.04	1	5000	0	EAEYFELPELVR
747.88	855.49	25.3	32.04	37.04	1	5000	0	EAEYFELPELVR
747.88	726.45	25.3	32.04	37.04	1	5000	0	EAEYFELPELVR
752.88	1012.57	25.3	32.04	37.04	1	5000	0	EAEYFELPELVR
752.88	865.50	25.3	32.04	37.04	1	5000	0	EAEYFELPELVR
752.88	736.46	25.3	32.04	37.04	1	5000	0	EAEYFELPELVR
629.81	997.50	21.8	31.03	36.03	1	5000	0	FNFLEQAFDK
629.81	850.43	21.8	31.03	36.03	1	5000	0	FNFLEQAFDK
629.81	737.35	21.8	31.03	36.03	1	5000	0	FNFLEQAFDK
633.82	1005.51	21.8	31.03	36.03	1	5000	0	FNFLEQAFDK
633.82	858.44	21.8	31.03	36.03	1	5000	0	FNFLEQAFDK
633.82	745.36	21.8	31.03	36.03	1	5000	0	FNFLEQAFDK
533.72	865.37	18.9	23.84	28.84	1	5000	0	SNC[+57.0]YGYFR

Supplementary table 2 contd.

533.72	705.34	18.9	23.84	28.84	1	5000	0	SNC[+57.0]YGYFR
533.72	542.27	18.9	23.84	28.84	1	5000	0	SNC[+57.0]YGYFR
538.73	875.37	18.9	23.84	28.84	1	5000	0	SNC[+57.0]YGYFR
538.73	715.34	18.9	23.84	28.84	1	5000	0	SNC[+57.0]YGYFR
538.73	552.28	18.9	23.84	28.84	1	5000	0	SNC[+57.0]YGYFR
794.39	1086.52	26.7	29.1	34.1	1	5000	0	EASTIAEYISGYQR
794.39	886.44	26.7	29.1	34.1	1	5000	0	EASTIAEYISGYQR
794.39	610.29	26.7	29.1	34.1	1	5000	0	EASTIAEYISGYQR
799.39	1096.53	26.7	29.1	34.1	1	5000	0	EASTIAEYISGYQR
799.39	896.45	26.7	29.1	34.1	1	5000	0	EASTIAEYISGYQR
799.39	620.30	26.7	29.1	34.1	1	5000	0	EASTIAEYISGYQR
784.10	1225.65	32.1	30.67	35.67	1	5000	0	LITINQQWKPIEELQNVQR
784.10	886.47	32.1	30.67	35.67	1	5000	0	LITINQQWKPIEELQNVQR
787.44	1235.66	32.1	30.67	35.67	1	5000	0	LITINQQWKPIEELQNVQR
787.44	896.48	32.1	30.67	35.67	1	5000	0	LITINQQWKPIEELQNVQR
416.23	730.41	15.4	21.82	26.82	1	5000	0	TDVILDR
416.23	615.38	15.4	21.82	26.82	1	5000	0	TDVILDR
421.24	740.42	15.4	21.82	26.82	1	5000	0	TDVILDR
421.24	625.39	15.4	21.82	26.82	1	5000	0	TDVILDR
613.32	1166.66	25.6	27.68	32.68	1	5000	0	NVTGHYISPFHDIPLK
613.32	1053.57	25.6	27.68	32.68	1	5000	0	NVTGHYISPFHDIPLK
613.32	966.54	25.6	27.68	32.68	1	5000	0	NVTGHYISPFHDIPLK
616.00	1174.67	25.6	27.68	32.68	1	5000	0	NVTGHYISPFHDIPLK
616.00	1061.59	25.6	27.68	32.68	1	5000	0	NVTGHYISPFHDIPLK
616.00	974.55	25.6	27.68	32.68	1	5000	0	NVTGHYISPFHDIPLK
678.36	1129.55	23.3	23.93	28.93	1	5000	0	LIAINANDPEASK
678.36	945.43	23.3	23.93	28.93	1	5000	0	LIAINANDPEASK
678.36	531.28	23.3	23.93	28.93	1	5000	0	LIAINANDPEASK
682.37	1137.56	23.3	23.93	28.93	1	5000	0	LIAINANDPEASK
682.37	953.44	23.3	23.93	28.93	1	5000	0	LIAINANDPEASK
682.37	539.29	23.3	23.93	28.93	1	5000	0	LIAINANDPEASK
445.77	819.50	16.3	27.77	32.77	1	5000	0	AFALEVIK
445.77	672.43	16.3	27.77	32.77	1	5000	0	AFALEVIK
445.77	601.39	16.3	27.77	32.77	1	5000	0	AFALEVIK
449.78	827.51	16.3	27.77	32.77	1	5000	0	AFALEVIK
449.78	680.44	16.3	27.77	32.77	1	5000	0	AFALEVIK
449.78	609.41	16.3	27.77	32.77	1	5000	0	AFALEVIK
471.24	741.36	17	26.68	31.68	1	5000	0	SLDNFFAK
471.24	626.33	17	26.68	31.68	1	5000	0	SLDNFFAK
471.24	512.29	17	26.68	31.68	1	5000	0	SLDNFFAK
475.25	749.37	17	26.68	31.68	1	5000	0	SLDNFFAK
475.25	634.34	17	26.68	31.68	1	5000	0	SLDNFFAK
475.25	520.30	17	26.68	31.68	1	5000	0	SLDNFFAK
706.40	1042.59	24.1	26.62	31.62	1	5000	0	GALQNIIPASTGAAK
706.40	928.55	24.1	26.62	31.62	1	5000	0	GALQNIIPASTGAAK
706.40	815.46	24.1	26.62	31.62	1	5000	0	GALQNIIPASTGAAK
710.41	1050.60	24.1	26.62	31.62	1	5000	0	GALQNIIPASTGAAK
710.41	936.56	24.1	26.62	31.62	1	5000	0	GALQNIIPASTGAAK
710.41	823.48	24.1	26.62	31.62	1	5000	0	GALQNIIPASTGAAK
882.40	1101.46	29.4	29.84	34.84	1	5000	0	LISWYDNEFGYSNR
882.40	986.43	29.4	29.84	34.84	1	5000	0	LISWYDNEFGYSNR
882.40	743.35	29.4	29.84	34.84	1	5000	0	LISWYDNEFGYSNR
887.41	1111.47	29.4	29.84	34.84	1	5000	0	LISWYDNEFGYSNR
887.41	996.44	29.4	29.84	34.84	1	5000	0	LISWYDNEFGYSNR
887.41	753.36	29.4	29.84	34.84	1	5000	0	LISWYDNEFGYSNR
596.32	831.49	24.9	27.32	32.32	1	5000	0	HLEINPDHSIIETLR
596.32	744.46	24.9	27.32	32.32	1	5000	0	HLEINPDHSIIETLR
596.32	631.38	24.9	27.32	32.32	1	5000	0	HLEINPDHSIIETLR
599.66	841.50	24.9	27.32	32.32	1	5000	0	HLEINPDHSIIETLR
599.66	754.47	24.9	27.32	32.32	1	5000	0	HLEINPDHSIIETLR
599.66	641.39	24.9	27.32	32.32	1	5000	0	HLEINPDHSIIETLR
618.30	992.52	21.5	22.5	27.5	1	5000	0	DQVANSAFVER
618.30	893.45	21.5	22.5	27.5	1	5000	0	DQVANSAFVER
623.31	1002.52	21.5	22.5	27.5	1	5000	0	DQVANSAFVER
623.31	903.46	21.5	22.5	27.5	1	5000	0	DQVANSAFVER
891.42	1144.58	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
891.42	1057.55	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
891.42	960.50	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK

Supplementary table 2 contd.

891.42	774.40	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
891.42	659.37	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
895.42	1152.60	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
895.42	1065.57	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
895.42	968.51	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
895.42	782.41	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
895.42	667.39	29.6	28.12	33.12	1	5000	0	DYPDFSPSVDAEAIQK
721.36	1138.60	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
721.36	1051.57	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
721.36	994.55	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
721.36	879.52	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
721.36	587.41	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
721.36	474.33	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
725.36	1146.61	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
725.36	1059.58	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
725.36	1002.56	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
725.36	887.53	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
725.36	595.43	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
725.36	482.34	24.5	27.78	32.78	1	5000	0	SDTSGDYEITLLK
472.23	830.37	17.1	24.07	29.07	1	5000	0	LTFDEYR
472.23	729.32	17.1	24.07	29.07	1	5000	0	LTFDEYR
472.23	582.25	17.1	24.07	29.07	1	5000	0	LTFDEYR
472.23	467.22	17.1	24.07	29.07	1	5000	0	LTFDEYR
472.23	338.18	17.1	24.07	29.07	1	5000	0	LTFDEYR
477.23	840.38	17.1	24.07	29.07	1	5000	0	LTFDEYR
477.23	739.33	17.1	24.07	29.07	1	5000	0	LTFDEYR
477.23	592.26	17.1	24.07	29.07	1	5000	0	LTFDEYR
477.23	477.23	17.1	24.07	29.07	1	5000	0	LTFDEYR
477.23	348.19	17.1	24.07	29.07	1	5000	0	LTFDEYR

Supplementary table 3. Input file for SpectroDrive analysis (complete transition list).

ProteinId	StrippedSequence	iRT	Precursor Charge	Precursor Mz	Fragment Charge	Fragment Type	Fragment Number	Fragment Mz	Relative Fragment Intensity	ModifiedSequence	Workflow
sp|P00338|LDHA_HUMAN	DQLIYNLLK	82.81	2	560.32	1	y	6	763.47	750100.00	DQLIYNLLK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DQLIYNLLK	82.81	2	560.32	1	y	5	650.39	914700.00	DQLIYNLLK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DQLIYNLLK	82.81	2	560.32	1	y	4	487.32	262800.00	DQLIYNLLK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DQLIYNLLK	82.81	2	564.33	1	y	6	771.49	750100.00	DQLIYNLLK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DQLIYNLLK	82.81	2	564.33	1	y	5	658.40	914700.00	DQLIYNLLK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DQLIYNLLK	82.81	2	564.33	1	y	4	495.34	262800.00	DQLIYNLLK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DLADELALVDVIEDK	117.88	2	829.43	1	y	9	1001.55	194700.00	DLADELALVDVIEDK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DLADELALVDVIEDK	117.88	2	829.43	1	y	8	930.51	135400.00	DLADELALVDVIEDK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DLADELALVDVIEDK	117.88	2	829.43	1	y	6	718.36	131900.00	DLADELALVDVIEDK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DLADELALVDVIEDK	117.88	2	833.44	1	y	9	1009.57	194700.00	DLADELALVDVIEDK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DLADELALVDVIEDK	117.88	2	833.44	1	y	8	938.53	135400.00	DLADELALVDVIEDK	SPIKE_IN
sp|P00338|LDHA_HUMAN	DLADELALVDVIEDK	117.88	2	833.44	1	y	6	726.38	131900.00	DLADELALVDVIEDK	SPIKE_IN
sp|P00338|LDHA_HUMAN	QVVESAYEVIK	34.32	2	632.84	1	y	9	1037.55	14350.00	QVVESAYEVIK	SPIKE_IN
sp|P00338|LDHA_HUMAN	QVVESAYEVIK	34.32	2	632.84	1	y	8	938.48	13220.00	QVVESAYEVIK	SPIKE_IN
sp|P00338|LDHA_HUMAN	QVVESAYEVIK	34.32	2	632.84	1	y	7	809.44	11420.00	QVVESAYEVIK	SPIKE_IN
sp|P00338|LDHA_HUMAN	QVVESAYEVIK	34.32	2	636.85	1	y	9	1045.57	14350.00	QVVESAYEVIK	SPIKE_IN
sp|P00338|LDHA_HUMAN	QVVESAYEVIK	34.32	2	636.85	1	y	8	946.50	13220.00	QVVESAYEVIK	SPIKE_IN
sp|P00338|LDHA_HUMAN	QVVESAYEVIK	34.32	2	636.85	1	y	7	817.45	11420.00	QVVESAYEVIK	SPIKE_IN
sp|P04406|G3P_HUMAN	GALQNIIPASTGAAK	44.65	2	706.40	1	y	11	1042.59	971600.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN	GALQNIIPASTGAAK	44.65	2	706.40	1	y	10	928.55	750800.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN	GALQNIIPASTGAAK	44.65	2	706.40	1	y	9	815.46	5160000.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN	GALQNIIPASTGAAK	44.65	2	710.41	1	y	11	1050.60	971600.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN	GALQNIIPASTGAAK	44.65	2	710.41	1	y	10	936.56	750800.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN	GALQNIIPASTGAAK	44.65	2	710.41	1	y	9	823.48	5160000.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN	LISWYDNEFGYSNR	72.02	2	882.40	1	y	9	1101.46	177100.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN	LISWYDNEFGYSNR	72.02	2	882.40	1	y	8	986.43	79510.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN	LISWYDNEFGYSNR	72.02	2	882.40	1	y	6	743.35	152300.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN	LISWYDNEFGYSNR	72.02	2	887.41	1	y	9	1111.47	177100.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN	LISWYDNEFGYSNR	72.02	2	887.41	1	y	8	996.44	79510.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN	LISWYDNEFGYSNR	72.02	2	887.41	1	y	6	753.36	152300.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P07195|LDHB_HUMAN	LIAPVAEEEATVPNNK	33.47	2	847.95	1	y	11	1201.57	467700.00	LIAPVAEEEATVPNNK	SPIKE_IN
sp|P07195|LDHB_HUMAN	LIAPVAEEEATVPNNK	33.47	2	847.95	1	y	5	571.32	179900.00	LIAPVAEEEATVPNNK	SPIKE_IN
sp|P07195|LDHB_HUMAN	LIAPVAEEEATVPNNK	33.47	2	847.95	1	y	4	472.25	1845000.00	LIAPVAEEEATVPNNK	SPIKE_IN
sp|P07195|LDHB_HUMAN	LIAPVAEEEATVPNNK	33.47	2	851.96	1	y	11	1209.58	467700.00	LIAPVAEEEATVPNNK	SPIKE_IN

Supplementary table 3 contd.

sp|P07195|LDHB_HUMAN	LIAPVAEEEATVPNNK	33.47	2	851.96	1	y	5	579.33	179900.00	LIAPVAEEEATVPNNK	SPIKE_IN
sp|P07195|LDHB_HUMAN	LIAPVAEEEATVPNNK	33.47	2	851.96	1	y	4	480.27	1845000.00	LIAPVAEEEATVPNNK	SPIKE_IN
sp|P07195|LDHB_HUMAN	SLADELALVDVLEDK	116.93	2	815.43	1	y	9	1001.55	1069.00	SLADELALVDVLEDK	SPIKE_IN
sp|P07195|LDHB_HUMAN	SLADELALVDVLEDK	116.93	2	815.43	1	y	8	930.51	698.90	SLADELALVDVLEDK	SPIKE_IN
sp|P07195|LDHB_HUMAN	SLADELALVDVLEDK	116.93	2	815.43	1	y	7	817.43	673.20	SLADELALVDVLEDK	SPIKE_IN
sp|P07195|LDHB_HUMAN	SLADELALVDVLEDK	116.93	2	819.44	1	y	9	1009.57	1069.00	SLADELALVDVLEDK	SPIKE_IN
sp|P07195|LDHB_HUMAN	SLADELALVDVLEDK	116.93	2	819.44	1	y	8	938.53	698.90	SLADELALVDVLEDK	SPIKE_IN
sp|P07195|LDHB_HUMAN	SLADELALVDVLEDK	116.93	2	819.44	1	y	7	825.44	673.20	SLADELALVDVLEDK	SPIKE_IN
sp|P07195|LDHB_HUMAN	IVVVTAGVR	18.66	2	457.30	1	y	7	701.43	2742000.00	IVVVTAGVR	SPIKE_IN
sp|P07195|LDHB_HUMAN	IVVVTAGVR	18.66	2	457.30	1	y	6	602.36	1159000.00	IVVVTAGVR	SPIKE_IN
sp|P07195|LDHB_HUMAN	IVVVTAGVR	18.66	2	457.30	1	y	5	503.29	820000.00	IVVVTAGVR	SPIKE_IN
sp|P07195|LDHB_HUMAN	IVVVTAGVR	18.66	2	462.30	1	y	7	711.44	2742000.00	IVVVTAGVR	SPIKE_IN
sp|P07195|LDHB_HUMAN	IVVVTAGVR	18.66	2	462.30	1	y	6	612.37	1159000.00	IVVVTAGVR	SPIKE_IN
sp|P07195|LDHB_HUMAN	IVVVTAGVR	18.66	2	462.30	1	y	5	513.30	820000.00	IVVVTAGVR	SPIKE_IN
sp|P07900|HS90A_HUMAN	HLEINPDHSIIETLR	50.57	3	596.32	1	y	7	831.49	1195000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN	HLEINPDHSIIETLR	50.57	3	596.32	1	y	6	744.46	531500.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN	HLEINPDHSIIETLR	50.57	3	596.32	1	y	5	631.38	624000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN	HLEINPDHSIIETLR	50.57	3	599.66	1	y	7	841.50	1195000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN	HLEINPDHSIIETLR	50.57	3	599.66	1	y	6	754.47	531500.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN	HLEINPDHSIIETLR	50.57	3	599.66	1	y	5	641.39	624000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN	DQVANSAFVER	9.64	2	618.30	1	y	9	992.52	461800.00	DQVANSAFVER	SPIKE_IN
sp|P07900|HS90A_HUMAN	DQVANSAFVER	9.64	2	618.30	1	y	8	893.45	775100.00	DQVANSAFVER	SPIKE_IN
sp|P07900|HS90A_HUMAN	DQVANSAFVER	9.64	2	623.31	1	y	9	1002.52	461800.00	DQVANSAFVER	SPIKE_IN
sp|P07900|HS90A_HUMAN	DQVANSAFVER	9.64	2	623.31	1	y	8	903.46	775100.00	DQVANSAFVER	SPIKE_IN
sp|P11413|G6PD_HUMAN	NSYVAGQYDDAASYQR	19.81	2	904.40	1	y	12	1344.58	8702.00	NSYVAGQYDDAASYQR	SPIKE_IN
sp|P11413|G6PD_HUMAN	NSYVAGQYDDAASYQR	19.81	2	904.40	1	y	9	1088.46	8640.00	NSYVAGQYDDAASYQR	SPIKE_IN
sp|P11413|G6PD_HUMAN	NSYVAGQYDDAASYQR	19.81	2	904.40	1	y	8	925.40	8096.00	NSYVAGQYDDAASYQR	SPIKE_IN
sp|P11413|G6PD_HUMAN	NSYVAGQYDDAASYQR	19.81	2	909.40	1	y	12	1354.59	8702.00	NSYVAGQYDDAASYQR	SPIKE_IN
sp|P11413|G6PD_HUMAN	NSYVAGQYDDAASYQR	19.81	2	909.40	1	y	9	1098.47	8640.00	NSYVAGQYDDAASYQR	SPIKE_IN
sp|P11413|G6PD_HUMAN	NSYVAGQYDDAASYQR	19.81	2	909.40	1	y	8	935.41	8096.00	NSYVAGQYDDAASYQR	SPIKE_IN
sp|P11413|G6PD_HUMAN	GGYFDEFGIIR	81.28	2	637.31	1	y	8	996.51	265500.00	GGYFDEFGIIR	SPIKE_IN
sp|P11413|G6PD_HUMAN	GGYFDEFGIIR	81.28	2	637.31	1	y	7	849.45	389300.00	GGYFDEFGIIR	SPIKE_IN
sp|P11413|G6PD_HUMAN	GGYFDEFGIIR	81.28	2	637.31	1	y	5	605.38	271600.00	GGYFDEFGIIR	SPIKE_IN
sp|P11413|G6PD_HUMAN	GGYFDEFGIIR	81.28	2	642.32	1	y	8	1006.52	265500.00	GGYFDEFGIIR	SPIKE_IN
sp|P11413|G6PD_HUMAN	GGYFDEFGIIR	81.28	2	642.32	1	y	7	859.45	389300.00	GGYFDEFGIIR	SPIKE_IN
sp|P11413|G6PD_HUMAN	GGYFDEFGIIR	81.28	2	642.32	1	y	5	615.39	271600.00	GGYFDEFGIIR	SPIKE_IN

Supplementary table 3 contd.

sp|P12429|ANXA3_HUMAN	DYPDFSPSVDAEAIQK	57.41	2	891.42	1	y	11	1144.58	358100.00	DYPDFSPSVDAEAIQK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	DYPDFSPSVDAEAIQK	57.41	2	891.42	1	y	10	1057.55	700900.00	DYPDFSPSVDAEAIQK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	DYPDFSPSVDAEAIQK	57.41	2	895.42	1	y	11	1152.60	358100.00	DYPDFSPSVDAEAIQK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	DYPDFSPSVDAEAIQK	57.41	2	895.42	1	y	10	1065.57	700900.00	DYPDFSPSVDAEAIQK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	SDTSGDYEITLLK	54.51	2	721.36	1	y	10	1138.60	13560.00	SDTSGDYEITLLK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	SDTSGDYEITLLK	54.51	2	721.36	1	y	9	1051.57	24290.00	SDTSGDYEITLLK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	SDTSGDYEITLLK	54.51	2	725.36	1	y	10	1146.61	13560.00	SDTSGDYEITLLK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	SDTSGDYEITLLK	54.51	2	725.36	1	y	9	1059.58	24290.00	SDTSGDYEITLLK	SPIKE_IN
sp|P12429|ANXA3_HUMAN	LTFDEYR	22.99	2	472.23	1	y	5	729.32	3802000.00	LTFDEYR	SPIKE_IN
sp|P12429|ANXA3_HUMAN	LTFDEYR	22.99	2	472.23	1	y	4	582.25	493000.00	LTFDEYR	SPIKE_IN
sp|P12429|ANXA3_HUMAN	LTFDEYR	22.99	2	477.23	1	y	5	739.33	3802000.00	LTFDEYR	SPIKE_IN
sp|P12429|ANXA3_HUMAN	LTFDEYR	22.99	2	477.23	1	y	4	592.26	493000.00	LTFDEYR	SPIKE_IN
sp|P15121|ALDR_HUMAN	SPPGQVTEAVK	-9.56	2	556.80	1	y	9	928.51	442800.00	SPPGQVTEAVK	SPIKE_IN
sp|P15121|ALDR_HUMAN	SPPGQVTEAVK	-9.56	2	556.80	1	y	8	831.46	53200.00	SPPGQVTEAVK	SPIKE_IN
sp|P15121|ALDR_HUMAN	SPPGQVTEAVK	-9.56	2	556.80	1	y	6	646.38	34170.00	SPPGQVTEAVK	SPIKE_IN
sp|P15121|ALDR_HUMAN	SPPGQVTEAVK	-9.56	2	560.81	1	y	9	936.52	442800.00	SPPGQVTEAVK	SPIKE_IN
sp|P15121|ALDR_HUMAN	SPPGQVTEAVK	-9.56	2	560.81	1	y	8	839.47	53200.00	SPPGQVTEAVK	SPIKE_IN
sp|P15121|ALDR_HUMAN	SPPGQVTEAVK	-9.56	2	560.81	1	y	6	654.39	34170.00	SPPGQVTEAVK	SPIKE_IN
sp|P15121|ALDR_HUMAN	MPILGLGTWK	81.78	2	558.32	1	y	8	887.53	13140.00	MPILGLGTWK	SPIKE_IN
sp|P15121|ALDR_HUMAN	MPILGLGTWK	81.78	2	558.32	1	y	7	774.45	21660.00	MPILGLGTWK	SPIKE_IN
sp|P15121|ALDR_HUMAN	MPILGLGTWK	81.78	2	558.32	1	y	6	661.37	29920.00	MPILGLGTWK	SPIKE_IN
sp|P15121|ALDR_HUMAN	MPILGLGTWK	81.78	2	562.32	1	y	8	895.55	13140.00	MPILGLGTWK	SPIKE_IN
sp|P15121|ALDR_HUMAN	MPILGLGTWK	81.78	2	562.32	1	y	7	782.47	21660.00	MPILGLGTWK	SPIKE_IN
sp|P15121|ALDR_HUMAN	MPILGLGTWK	81.78	2	562.32	1	y	6	669.38	29920.00	MPILGLGTWK	SPIKE_IN
sp|P15121|ALDR_HUMAN	VAIDVGYR	19.56	2	446.75	1	y	6	722.38	235200.00	VAIDVGYR	SPIKE_IN
sp|P15121|ALDR_HUMAN	VAIDVGYR	19.56	2	446.75	1	y	5	609.30	140900.00	VAIDVGYR	SPIKE_IN
sp|P15121|ALDR_HUMAN	VAIDVGYR	19.56	2	446.75	1	y	4	494.27	45000.00	VAIDVGYR	SPIKE_IN
sp|P15121|ALDR_HUMAN	VAIDVGYR	19.56	2	451.75	1	y	6	732.39	235200.00	VAIDVGYR	SPIKE_IN
sp|P15121|ALDR_HUMAN	VAIDVGYR	19.56	2	451.75	1	y	5	619.31	140900.00	VAIDVGYR	SPIKE_IN
sp|P15121|ALDR_HUMAN	VAIDVGYR	19.56	2	451.75	1	y	4	504.28	45000.00	VAIDVGYR	SPIKE_IN
sp|P19338|NUCL_HUMAN	FGYVDFESAEDLEK	73.18	2	824.87	1	y	10	1182.52	1537.00	FGYVDFESAEDLEK	SPIKE_IN
sp|P19338|NUCL_HUMAN	FGYVDFESAEDLEK	73.18	2	824.87	1	y	9	1067.49	843.20	FGYVDFESAEDLEK	SPIKE_IN
sp|P19338|NUCL_HUMAN	FGYVDFESAEDLEK	73.18	2	824.87	1	y	7	791.38	961.40	FGYVDFESAEDLEK	SPIKE_IN
sp|P19338|NUCL_HUMAN	FGYVDFESAEDLEK	73.18	2	828.88	1	y	10	1190.53	1537.00	FGYVDFESAEDLEK	SPIKE_IN
sp|P19338|NUCL_HUMAN	FGYVDFESAEDLEK	73.18	2	828.88	1	y	9	1075.50	843.20	FGYVDFESAEDLEK	SPIKE_IN
sp|P19338|NUCL_HUMAN	FGYVDFESAEDLEK	73.18	2	828.88	1	y	7	799.39	961.40	FGYVDFESAEDLEK	SPIKE_IN

Supplementary table 3 contd.

sp|P19338|NUCL_HUMAN	EVFEDAAEIR	31.54	2	589.79	1	y	8	950.46	51910.00	EVFEDAAEIR	SPIKE_IN
sp|P19338|NUCL_HUMAN	EVFEDAAEIR	31.54	2	589.79	1	y	7	803.39	32050.00	EVFEDAAEIR	SPIKE_IN
sp|P19338|NUCL_HUMAN	EVFEDAAEIR	31.54	2	589.79	1	y	6	674.35	31030.00	EVFEDAAEIR	SPIKE_IN
sp|P19338|NUCL_HUMAN	EVFEDAAEIR	31.54	2	594.79	1	y	8	960.47	51910.00	EVFEDAAEIR	SPIKE_IN
sp|P19338|NUCL_HUMAN	EVFEDAAEIR	31.54	2	594.79	1	y	7	813.40	32050.00	EVFEDAAEIR	SPIKE_IN
sp|P19338|NUCL_HUMAN	EVFEDAAEIR	31.54	2	594.79	1	y	6	684.36	31030.00	EVFEDAAEIR	SPIKE_IN
sp|P19338|NUCL_HUMAN	GFGFVDFNSEEDAK	69.63	2	781.34	1	y	10	1153.50	438000.00	GFGFVDFNSEEDAK	SPIKE_IN
sp|P19338|NUCL_HUMAN	GFGFVDFNSEEDAK	69.63	2	781.34	1	y	9	1054.43	738300.00	GFGFVDFNSEEDAK	SPIKE_IN
sp|P19338|NUCL_HUMAN	GFGFVDFNSEEDAK	69.63	2	781.34	1	y	8	939.41	332100.00	GFGFVDFNSEEDAK	SPIKE_IN
sp|P19338|NUCL_HUMAN	GFGFVDFNSEEDAK	69.63	2	785.35	1	y	10	1161.51	438000.00	GFGFVDFNSEEDAK	SPIKE_IN
sp|P19338|NUCL_HUMAN	GFGFVDFNSEEDAK	69.63	2	785.35	1	y	9	1062.45	738300.00	GFGFVDFNSEEDAK	SPIKE_IN
sp|P19338|NUCL_HUMAN	GFGFVDFNSEEDAK	69.63	2	785.35	1	y	8	947.42	332100.00	GFGFVDFNSEEDAK	SPIKE_IN
sp|P22223|CADH3_HUMAN	STGTISVISSGLDR	43.28	2	696.87	1	y	9	933.50	2257000.00	STGTISVISSGLDR	SPIKE_IN
sp|P22223|CADH3_HUMAN	STGTISVISSGLDR	43.28	2	696.87	1	y	7	747.40	834400.00	STGTISVISSGLDR	SPIKE_IN
sp|P22223|CADH3_HUMAN	STGTISVISSGLDR	43.28	2	696.87	1	y	6	634.32	480500.00	STGTISVISSGLDR	SPIKE_IN
sp|P22223|CADH3_HUMAN	STGTISVISSGLDR	43.28	2	701.87	1	y	9	943.51	2257000.00	STGTISVISSGLDR	SPIKE_IN
sp|P22223|CADH3_HUMAN	STGTISVISSGLDR	43.28	2	701.87	1	y	7	757.41	834400.00	STGTISVISSGLDR	SPIKE_IN
sp|P22223|CADH3_HUMAN	STGTISVISSGLDR	43.28	2	701.87	1	y	6	644.32	480500.00	STGTISVISSGLDR	SPIKE_IN
sp|P22223|CADH3_HUMAN	LTVTDLDAPNSPAWR	59.46	2	828.42	1	y	9	1013.48	1537000.00	LTVTDLDAPNSPAWR	SPIKE_IN
sp|P22223|CADH3_HUMAN	LTVTDLDAPNSPAWR	59.46	2	828.42	1	y	8	898.45	2471000.00	LTVTDLDAPNSPAWR	SPIKE_IN
sp|P22223|CADH3_HUMAN	LTVTDLDAPNSPAWR	59.46	2	828.42	1	y	7	827.42	2043000.00	LTVTDLDAPNSPAWR	SPIKE_IN
sp|P22223|CADH3_HUMAN	LTVTDLDAPNSPAWR	59.46	2	833.43	1	y	9	1023.49	1537000.00	LTVTDLDAPNSPAWR	SPIKE_IN
sp|P22223|CADH3_HUMAN	LTVTDLDAPNSPAWR	59.46	2	833.43	1	y	8	908.46	2471000.00	LTVTDLDAPNSPAWR	SPIKE_IN
sp|P22223|CADH3_HUMAN	LTVTDLDAPNSPAWR	59.46	2	833.43	1	y	7	837.42	2043000.00	LTVTDLDAPNSPAWR	SPIKE_IN
sp|P27695|APEX1_HUMAN	EAAGEGPALYEDPPDQK	16.82	2	893.91	1	y	12	1329.63	131400.00	EAAGEGPALYEDPPDQK	SPIKE_IN
sp|P27695|APEX1_HUMAN	EAAGEGPALYEDPPDQK	16.82	2	893.91	1	y	8	991.44	224000.00	EAAGEGPALYEDPPDQK	SPIKE_IN
sp|P27695|APEX1_HUMAN	EAAGEGPALYEDPPDQK	16.82	2	893.91	1	y	5	584.30	2859000.00	EAAGEGPALYEDPPDQK	SPIKE_IN
sp|P27695|APEX1_HUMAN	EAAGEGPALYEDPPDQK	16.82	2	897.92	1	y	12	1337.65	131400.00	EAAGEGPALYEDPPDQK	SPIKE_IN
sp|P27695|APEX1_HUMAN	EAAGEGPALYEDPPDQK	16.82	2	897.92	1	y	8	999.45	224000.00	EAAGEGPALYEDPPDQK	SPIKE_IN
sp|P27695|APEX1_HUMAN	EAAGEGPALYEDPPDQK	16.82	2	897.92	1	y	5	592.32	2859000.00	EAAGEGPALYEDPPDQK	SPIKE_IN
sp|P27695|APEX1_HUMAN	QGFGELLQAVPLADSFR	119.43	2	924.49	1	y	10	1103.58	1906.00	QGFGELLQAVPLADSFR	SPIKE_IN
sp|P27695|APEX1_HUMAN	QGFGELLQAVPLADSFR	119.43	2	924.49	1	y	9	975.53	1749.00	QGFGELLQAVPLADSFR	SPIKE_IN
sp|P27695|APEX1_HUMAN	QGFGELLQAVPLADSFR	119.43	2	924.49	1	y	7	805.42	3167.00	QGFGELLQAVPLADSFR	SPIKE_IN
sp|P27695|APEX1_HUMAN	QGFGELLQAVPLADSFR	119.43	2	929.49	1	y	10	1113.59	1906.00	QGFGELLQAVPLADSFR	SPIKE_IN
sp|P27695|APEX1_HUMAN	QGFGELLQAVPLADSFR	119.43	2	929.49	1	y	9	985.53	1749.00	QGFGELLQAVPLADSFR	SPIKE_IN
sp|P27695|APEX1_HUMAN	QGFGELLQAVPLADSFR	119.43	2	929.49	1	y	7	815.43	3167.00	QGFGELLQAVPLADSFR	SPIKE_IN

Supplementary table 3 contd.

sp|P27695|APEX1_HUMAN	LDYFLLSHSLLPALCDSK	113.98	2	1046.54	1	y	7	790.38	88430.00	LDYFLLSHSLLPALC[+57]DSK	SPIKE_IN
sp|P27695|APEX1_HUMAN	LDYFLLSHSLLPALCDSK	113.98	2	1046.54	1	y	5	622.29	22500.00	LDYFLLSHSLLPALC[+57]DSK	SPIKE_IN
sp|P27695|APEX1_HUMAN	LDYFLLSHSLLPALCDSK	113.98	2	1046.54	1	b	3	392.18	31790.00	LDYFLLSHSLLPALC[+57]DSK	SPIKE_IN
sp|P27695|APEX1_HUMAN	LDYFLLSHSLLPALCDSK	113.98	2	1050.55	1	y	7	798.39	88430.00	LDYFLLSHSLLPALC[+57]DSK	SPIKE_IN
sp|P27695|APEX1_HUMAN	LDYFLLSHSLLPALCDSK	113.98	2	1050.55	1	y	5	630.30	22500.00	LDYFLLSHSLLPALC[+57]DSK	SPIKE_IN
sp|P27695|APEX1_HUMAN	LDYFLLSHSLLPALCDSK	113.98	2	1050.55	1	b	3	392.18	31790.00	LDYFLLSHSLLPALC[+57]DSK	SPIKE_IN
sp|P31949|S10AB_HUMAN	ISSPTETER	-8.63	2	510.25	1	y	8	906.42	6767000.00	ISSPTETER	SPIKE_IN
sp|P31949|S10AB_HUMAN	ISSPTETER	-8.63	2	510.25	1	y	7	819.38	3382000.00	ISSPTETER	SPIKE_IN
sp|P31949|S10AB_HUMAN	ISSPTETER	-8.63	2	510.25	1	y	6	732.35	3798000.00	ISSPTETER	SPIKE_IN
sp|P31949|S10AB_HUMAN	ISSPTETER	-8.63	2	515.26	1	y	8	916.42	6767000.00	ISSPTETER	SPIKE_IN
sp|P31949|S10AB_HUMAN	ISSPTETER	-8.63	2	515.26	1	y	7	829.39	3382000.00	ISSPTETER	SPIKE_IN
sp|P31949|S10AB_HUMAN	ISSPTETER	-8.63	2	515.26	1	y	6	742.36	3798000.00	ISSPTETER	SPIKE_IN
sp|P31949|S10AB_HUMAN	DGYNYTLSK	11.1	2	530.75	1	y	6	725.38	1117000.00	DGYNYTLSK	SPIKE_IN
sp|P31949|S10AB_HUMAN	DGYNYTLSK	11.1	2	530.75	1	y	5	611.34	601200.00	DGYNYTLSK	SPIKE_IN
sp|P31949|S10AB_HUMAN	DGYNYTLSK	11.1	2	530.75	1	y	4	448.28	537800.00	DGYNYTLSK	SPIKE_IN
sp|P31949|S10AB_HUMAN	DGYNYTLSK	11.1	2	534.76	1	y	6	733.40	1117000.00	DGYNYTLSK	SPIKE_IN
sp|P31949|S10AB_HUMAN	DGYNYTLSK	11.1	2	534.76	1	y	5	619.35	601200.00	DGYNYTLSK	SPIKE_IN
sp|P31949|S10AB_HUMAN	DGYNYTLSK	11.1	2	534.76	1	y	4	456.29	537800.00	DGYNYTLSK	SPIKE_IN
sp|P31949|S10AB_HUMAN	DPGVLDR	-4.05	2	386.20	1	y	5	559.32	29260.00	DPGVLDR	SPIKE_IN
sp|P31949|S10AB_HUMAN	DPGVLDR	-4.05	2	386.20	1	y	4	502.30	5170.00	DPGVLDR	SPIKE_IN
sp|P31949|S10AB_HUMAN	DPGVLDR	-4.05	2	386.20	1	y	3	403.23	9090.00	DPGVLDR	SPIKE_IN
sp|P31949|S10AB_HUMAN	DPGVLDR	-4.05	2	391.21	1	y	5	569.33	29260.00	DPGVLDR	SPIKE_IN
sp|P31949|S10AB_HUMAN	DPGVLDR	-4.05	2	391.21	1	y	4	512.31	5170.00	DPGVLDR	SPIKE_IN
sp|P31949|S10AB_HUMAN	DPGVLDR	-4.05	2	391.21	1	y	3	413.24	9090.00	DPGVLDR	SPIKE_IN
sp|P36952|SPB5_HUMAN	IIELPFQNK	58.65	2	551.32	1	y	6	746.42	903400.00	IIELPFQNK	SPIKE_IN
sp|P36952|SPB5_HUMAN	IIELPFQNK	58.65	2	551.32	1	y	5	633.34	1793000.00	IIELPFQNK	SPIKE_IN
sp|P36952|SPB5_HUMAN	IIELPFQNK	58.65	2	555.33	1	y	6	754.43	903400.00	IIELPFQNK	SPIKE_IN
sp|P36952|SPB5_HUMAN	IIELPFQNK	58.65	2	555.33	1	y	5	641.35	1793000.00	IIELPFQNK	SPIKE_IN
sp|P39687|AN32A_HUMAN	ELVLDNSR	3.51	2	473.25	1	y	6	703.37	20500.00	ELVLDNSR	SPIKE_IN
sp|P39687|AN32A_HUMAN	ELVLDNSR	3.51	2	473.25	1	y	5	604.30	27810.00	ELVLDNSR	SPIKE_IN
sp|P39687|AN32A_HUMAN	ELVLDNSR	3.51	2	473.25	1	y	4	491.22	13890.00	ELVLDNSR	SPIKE_IN
sp|P39687|AN32A_HUMAN	ELVLDNSR	3.51	2	478.26	1	y	6	713.38	20500.00	ELVLDNSR	SPIKE_IN
sp|P39687|AN32A_HUMAN	ELVLDNSR	3.51	2	478.26	1	y	5	614.31	27810.00	ELVLDNSR	SPIKE_IN
sp|P39687|AN32A_HUMAN	ELVLDNSR	3.51	2	478.26	1	y	4	501.23	13890.00	ELVLDNSR	SPIKE_IN
sp|P39687|AN32A_HUMAN	CPNLTHLNLSGNK	25.69	3	489.92	1	y	7	745.42	468400.00	C[+57]PNLTHLNLSGNK	SPIKE_IN
sp|P39687|AN32A_HUMAN	CPNLTHLNLSGNK	25.69	3	489.92	1	y	6	632.34	478000.00	C[+57]PNLTHLNLSGNK	SPIKE_IN

Supplementary table 3 contd.

sp|P39687|AN32A_HUMAN	CPNLTHLNLSGNK	25.69	3	489.92	1	y	4	405.21	284700.00	C[+57]PNLTHLNLSGNK	SPIKE_IN
sp|P39687|AN32A_HUMAN	CPNLTHLNLSGNK	25.69	3	492.59	1	y	7	753.43	468400.00	C[+57]PNLTHLNLSGNK	SPIKE_IN
sp|P39687|AN32A_HUMAN	CPNLTHLNLSGNK	25.69	3	492.59	1	y	6	640.35	478000.00	C[+57]PNLTHLNLSGNK	SPIKE_IN
sp|P39687|AN32A_HUMAN	CPNLTHLNLSGNK	25.69	3	492.59	1	y	4	413.22	284700.00	C[+57]PNLTHLNLSGNK	SPIKE_IN
sp|P55011|S12A2_HUMAN	VELPGTAVPSVPEDAAPASR	46.04	2	982.01	1	y	13	1295.66	995200.00	VELPGTAVPSVPEDAAPASR	SPIKE_IN
sp|P55011|S12A2_HUMAN	VELPGTAVPSVPEDAAPASR	46.04	2	982.01	1	y	12	1196.59	7980000.00	VELPGTAVPSVPEDAAPASR	SPIKE_IN
sp|P55011|S12A2_HUMAN	VELPGTAVPSVPEDAAPASR	46.04	2	982.01	1	y	9	913.44	1397000.00	VELPGTAVPSVPEDAAPASR	SPIKE_IN
sp|P55011|S12A2_HUMAN	VELPGTAVPSVPEDAAPASR	46.04	2	987.01	1	y	13	1305.67	995200.00	VELPGTAVPSVPEDAAPASR	SPIKE_IN
sp|P55011|S12A2_HUMAN	VELPGTAVPSVPEDAAPASR	46.04	2	987.01	1	y	12	1206.60	7980000.00	VELPGTAVPSVPEDAAPASR	SPIKE_IN
sp|P55011|S12A2_HUMAN	VELPGTAVPSVPEDAAPASR	46.04	2	987.01	1	y	9	923.45	1397000.00	VELPGTAVPSVPEDAAPASR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	2	687.84	1	y	9	993.51	515000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	2	687.84	1	y	8	922.47	1686000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	2	687.84	1	y	6	726.35	522600.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	2	692.85	1	y	9	1003.52	515000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	2	692.85	1	y	8	932.48	1686000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	2	692.85	1	y	6	736.36	522600.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	3	458.90	1	y	8	922.47	3103000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	3	458.90	1	y	6	726.35	3745000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	3	458.90	1	y	5	589.29	2648000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	3	462.23	1	y	8	932.48	3103000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	3	462.23	1	y	6	736.36	3745000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	AFYAPVHADDLR	30.67	3	462.23	1	y	5	599.30	2648000.00	AFYAPVHADDLR	SPIKE_IN
sp|P55011|S12A2_HUMAN	EGLDISHLQGQEELLSSQEK	60.03	3	747.37	1	y	7	804.45	247300.00	EGLDISHLQGQEELLSSQEK	SPIKE_IN
sp|P55011|S12A2_HUMAN	EGLDISHLQGQEELLSSQEK	60.03	3	747.37	1	y	6	691.36	382500.00	EGLDISHLQGQEELLSSQEK	SPIKE_IN
sp|P55011|S12A2_HUMAN	EGLDISHLQGQEELLSSQEK	60.03	3	747.37	1	y	5	578.28	462100.00	EGLDISHLQGQEELLSSQEK	SPIKE_IN
sp|P55011|S12A2_HUMAN	EGLDISHLQGQEELLSSQEK	60.03	3	750.04	1	y	7	812.46	247300.00	EGLDISHLQGQEELLSSQEK	SPIKE_IN
sp|P55011|S12A2_HUMAN	EGLDISHLQGQEELLSSQEK	60.03	3	750.04	1	y	6	699.38	382500.00	EGLDISHLQGQEELLSSQEK	SPIKE_IN
sp|P55011|S12A2_HUMAN	EGLDISHLQGQEELLSSQEK	60.03	3	750.04	1	y	5	586.29	462100.00	EGLDISHLQGQEELLSSQEK	SPIKE_IN
sp|P55011|S12A2_HUMAN	ITDNELELYK	38.06	2	619.32	1	y	9	1124.55	20950.00	ITDNELELYK	SPIKE_IN
sp|P55011|S12A2_HUMAN	ITDNELELYK	38.06	2	619.32	1	y	8	1023.50	20530.00	ITDNELELYK	SPIKE_IN
sp|P55011|S12A2_HUMAN	ITDNELELYK	38.06	2	619.32	1	y	2	310.18	7514.00	ITDNELELYK	SPIKE_IN
sp|P55011|S12A2_HUMAN	ITDNELELYK	38.06	2	623.33	1	y	9	1132.56	20950.00	ITDNELELYK	SPIKE_IN
sp|P55011|S12A2_HUMAN	ITDNELELYK	38.06	2	623.33	1	y	8	1031.51	20530.00	ITDNELELYK	SPIKE_IN
sp|P55011|S12A2_HUMAN	ITDNELELYK	38.06	2	623.33	1	y	2	318.19	7514.00	ITDNELELYK	SPIKE_IN
sp|P62330|ARF6_HUMAN	FNVWDVGGQDK	53.03	2	632.80	1	y	9	1003.48	144000.00	FNVWDVGGQDK	SPIKE_IN
sp|P62330|ARF6_HUMAN	FNVWDVGGQDK	53.03	2	632.80	1	y	8	904.42	326500.00	FNVWDVGGQDK	SPIKE_IN

Supplementary table 3 contd.

sp|P62330|ARF6_HUMAN	FNVWDVGGQDK	53.03	2	632.80	1	y	7	718.34	145300.00	FNVWDVGGQDK	SPIKE_IN
sp|P62330|ARF6_HUMAN	FNVWDVGGQDK	53.03	2	636.81	1	y	9	1011.50	144000.00	FNVWDVGGQDK	SPIKE_IN
sp|P62330|ARF6_HUMAN	FNVWDVGGQDK	53.03	2	636.81	1	y	8	912.43	326500.00	FNVWDVGGQDK	SPIKE_IN
sp|P62330|ARF6_HUMAN	FNVWDVGGQDK	53.03	2	636.81	1	y	7	726.35	145300.00	FNVWDVGGQDK	SPIKE_IN
sp|P80188|NGAL_HUMAN	SYPGLTSYLVR	63.13	2	628.34	1	y	9	1005.57	189900.00	SYPGLTSYLVR	SPIKE_IN
sp|P80188|NGAL_HUMAN	SYPGLTSYLVR	63.13	2	628.34	1	y	8	908.52	30110.00	SYPGLTSYLVR	SPIKE_IN
sp|P80188|NGAL_HUMAN	SYPGLTSYLVR	63.13	2	633.34	1	y	9	1015.58	189900.00	SYPGLTSYLVR	SPIKE_IN
sp|P80188|NGAL_HUMAN	SYPGLTSYLVR	63.13	2	633.34	1	y	8	918.53	30110.00	SYPGLTSYLVR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	2	975.52	1	y	12	1333.75	4050.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	2	975.52	1	y	11	1220.66	43880.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	2	975.52	1	y	9	994.57	23530.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	2	980.52	1	y	12	1343.76	4050.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	2	980.52	1	y	11	1230.67	43880.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	2	980.52	1	y	9	1004.58	23530.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	3	650.68	1	y	9	994.57	789600.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	3	650.68	1	y	8	897.52	420900.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	3	650.68	1	y	7	840.49	200600.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	3	654.02	1	y	9	1004.58	789600.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	3	654.02	1	y	8	907.52	420900.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LENYPIPEPGPNEVLLR	75.82	3	654.02	1	y	7	850.50	200600.00	LENYPIPEPGPNEVLLR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	2	791.92	1	y	12	1226.64	141.40	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	2	791.92	1	y	11	1127.57	238.10	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	2	791.92	1	y	10	1014.49	250.20	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	2	795.92	1	y	12	1234.65	141.40	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	2	795.92	1	y	11	1135.58	238.10	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	2	795.92	1	y	10	1022.50	250.20	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	3	528.28	1	y	8	820.40	21700.00	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	3	528.28	1	y	7	691.36	31990.00	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	3	528.28	1	y	6	620.32	21030.00	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	3	530.95	1	y	8	828.42	21700.00	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	3	530.95	1	y	7	699.38	31990.00	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	KPMVLGHEASGTVEK	-9.53	3	530.95	1	y	6	628.34	21030.00	KPMVLGHEASGTVEK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LPDNVTFEEGALIEPLSVGIHACR	99.39	3	879.78	1	y	12	1351.72	11340.00	LPDNVTFEEGALIEPLSVGIHAC[+57]R	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LPDNVTFEEGALIEPLSVGIHACR	99.39	3	879.78	1	y	11	1238.63	21290.00	LPDNVTFEEGALIEPLSVGIHAC[+57]R	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LPDNVTFEEGALIEPLSVGIHACR	99.39	3	879.78	1	y	10	1109.59	53780.00	LPDNVTFEEGALIEPLSVGIHAC[+57]R	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LPDNVTFEEGALIEPLSVGIHACR	99.39	3	883.12	1	y	12	1361.72	11340.00	LPDNVTFEEGALIEPLSVGIHAC[+57]R	SPIKE_IN

Supplementary table 3 contd.

sp|Q00796|DHSO_HUMAN	LPDNVTFEEGALIEPLSVGIHACR	99.39	3	883.12	1	y	11	1248.64	21290.00	LPDNVTFEEGALIEPLSVGIHAC[+57]R	SPIKE_IN
sp|Q00796|DHSO_HUMAN	LPDNVTFEEGALIEPLSVGIHACR	99.39	3	883.12	1	y	10	1119.60	53780.00	LPDNVTFEEGALIEPLSVGIHAC[+57]R	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	2	795.42	1	y	9	961.53	3408.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	2	795.42	1	y	8	862.46	1728.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	2	795.42	1	y	7	763.39	1828.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	2	800.42	1	y	9	971.54	3408.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	2	800.42	1	y	8	872.47	1728.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	2	800.42	1	y	7	773.40	1828.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	3	530.62	1	y	4	434.24	31840.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	3	530.62	1	b	8	728.38	18620.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	3	530.62	1	b	11	1043.52	19460.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	3	533.95	1	y	4	444.24	31840.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	3	533.95	1	b	8	728.38	18620.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	AMGAAQVVVTDLSATR	64.42	3	533.95	1	b	11	1043.52	19460.00	AMGAAQVVVTDLSATR	SPIKE_IN
sp|Q00796|DHSO_HUMAN	EIGADLVLQISK	75.03	2	643.37	1	y	10	1043.61	487300.00	EIGADLVLQISK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	EIGADLVLQISK	75.03	2	643.37	1	y	6	687.44	239300.00	EIGADLVLQISK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	EIGADLVLQISK	75.03	2	643.37	1	y	5	588.37	367800.00	EIGADLVLQISK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	EIGADLVLQISK	75.03	2	647.38	1	y	10	1051.62	487300.00	EIGADLVLQISK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	EIGADLVLQISK	75.03	2	647.38	1	y	6	695.45	239300.00	EIGADLVLQISK	SPIKE_IN
sp|Q00796|DHSO_HUMAN	EIGADLVLQISK	75.03	2	647.38	1	y	5	596.39	367800.00	EIGADLVLQISK	SPIKE_IN
sp|Q01105|SET_HUMAN	VEVTEFEDIK	44.68	2	604.81	1	y	8	980.49	1824000.00	VEVTEFEDIK	SPIKE_IN
sp|Q01105|SET_HUMAN	VEVTEFEDIK	44.68	2	604.81	1	y	7	881.43	1900000.00	VEVTEFEDIK	SPIKE_IN
sp|Q01105|SET_HUMAN	VEVTEFEDIK	44.68	2	604.81	1	y	5	651.33	751300.00	VEVTEFEDIK	SPIKE_IN
sp|Q01105|SET_HUMAN	VEVTEFEDIK	44.68	2	608.81	1	y	8	988.51	1824000.00	VEVTEFEDIK	SPIKE_IN
sp|Q01105|SET_HUMAN	VEVTEFEDIK	44.68	2	608.81	1	y	7	889.44	1900000.00	VEVTEFEDIK	SPIKE_IN
sp|Q01105|SET_HUMAN	VEVTEFEDIK	44.68	2	608.81	1	y	5	659.35	751300.00	VEVTEFEDIK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	920.91	1	y	10	1302.56	887.10	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	920.91	1	y	9	1155.50	638.00	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	920.91	1	y	6	797.38	2297.00	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	920.91	1	y	4	537.27	1042.00	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	924.91	1	y	10	1310.58	887.10	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	924.91	1	y	9	1163.51	638.00	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	924.91	1	y	6	805.40	2297.00	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	IDFYFDENPYFENK	93.03	2	924.91	1	y	4	545.28	1042.00	IDFYFDENPYFENK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	2	723.83	1	y	10	1033.48	570.50	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	2	723.83	1	y	9	920.40	872.70	EFHLNESGDPSSK	SPIKE_IN

Supplementary table 3 contd.

sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	2	723.83	1	y	4	418.23	1670.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	2	727.84	1	y	10	1041.49	570.50	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	2	727.84	1	y	9	928.41	872.70	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	2	727.84	1	y	4	426.24	1670.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	3	482.89	1	y	6	590.28	845500.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	3	482.89	1	y	4	418.23	2314000.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	3	482.89	1	y	3	321.18	649300.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	3	485.56	1	y	6	598.29	845500.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	3	485.56	1	y	4	426.24	2314000.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q01105|SET_HUMAN	EFHLNESGDPSSK	2.16	3	485.56	1	y	3	329.19	649300.00	EFHLNESGDPSSK	SPIKE_IN
sp|Q15124|PGM5_HUMAN	VPVYETPAGWR	40.31	2	637.83	1	y	9	1078.53	41600.00	VPVYETPAGWR	SPIKE_IN
sp|Q15124|PGM5_HUMAN	VPVYETPAGWR	40.31	2	637.83	1	y	8	979.46	95580.00	VPVYETPAGWR	SPIKE_IN
sp|Q15124|PGM5_HUMAN	VPVYETPAGWR	40.31	2	637.83	1	y	7	816.40	56490.00	VPVYETPAGWR	SPIKE_IN
sp|Q15124|PGM5_HUMAN	VPVYETPAGWR	40.31	2	642.83	1	y	9	1088.54	41600.00	VPVYETPAGWR	SPIKE_IN
sp|Q15124|PGM5_HUMAN	VPVYETPAGWR	40.31	2	642.83	1	y	8	989.47	95580.00	VPVYETPAGWR	SPIKE_IN
sp|Q15124|PGM5_HUMAN	VPVYETPAGWR	40.31	2	642.83	1	y	7	826.41	56490.00	VPVYETPAGWR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	AAPFSLEYR	44.4	2	527.27	1	y	7	911.46	3517000.00	AAPFSLEYR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	AAPFSLEYR	44.4	2	527.27	1	y	6	814.41	9705000.00	AAPFSLEYR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	AAPFSLEYR	44.4	2	527.27	1	y	5	667.34	7233000.00	AAPFSLEYR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	AAPFSLEYR	44.4	2	532.28	1	y	7	921.47	3517000.00	AAPFSLEYR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	AAPFSLEYR	44.4	2	532.28	1	y	6	824.42	9705000.00	AAPFSLEYR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	AAPFSLEYR	44.4	2	532.28	1	y	5	677.35	7233000.00	AAPFSLEYR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	2	932.47	1	y	6	706.38	119300.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	2	932.47	1	y	4	496.24	70420.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	2	932.47	1	b	10	1158.56	62340.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	2	936.47	1	y	6	714.39	119300.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	2	936.47	1	y	4	504.25	70420.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	2	936.47	1	b	10	1158.56	62340.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	3	621.98	1	y	8	934.49	12680.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	3	621.98	1	y	6	706.38	25100.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	3	621.98	1	y	4	496.24	12050.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	3	624.65	1	y	8	942.50	12680.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	3	624.65	1	y	6	714.39	25100.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	GQYISPFHDIPIYADK	71.22	3	624.65	1	y	4	504.25	12050.00	GQYISPFHDIPIYADK	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	2	664.34	1	y	7	829.46	1759000.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	2	664.34	1	y	6	698.42	890200.00	DVFHMVVEVPR	SPIKE_IN

Supplementary table 3 contd.

sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	2	664.34	1	y	5	599.35	361600.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	2	669.35	1	y	7	839.47	1759000.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	2	669.35	1	y	6	708.43	890200.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	2	669.35	1	y	5	609.36	361600.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	3	443.23	1	y	7	829.46	38500.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	3	443.23	1	y	5	599.35	18900.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	3	443.23	1	y	4	500.28	27460.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	3	446.57	1	y	7	839.47	38500.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	3	446.57	1	y	5	609.36	18900.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15181|IPYR_HUMAN	DVFHMVVEVPR	70.25	3	446.57	1	y	4	510.29	27460.00	DVFHMVVEVPR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	650.65	1	b	2	324.15	24450.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	650.65	1	y	8	897.42	16190.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	650.65	1	y	7	782.39	14870.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	650.65	1	y	5	574.29	13250.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	653.98	1	b	2	324.15	24450.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	653.98	1	y	8	907.43	16190.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	653.98	1	y	7	792.40	14870.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	HWILPQDYDHAQAEAR	31.7	3	653.98	1	y	5	584.30	13250.00	HWILPQDYDHAQAEAR	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	485.26	1	y	7	806.44	5000.00	YIFDNVAK	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	485.26	1	y	6	693.36	377300.00	YIFDNVAK	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	485.26	1	y	5	546.29	28990.00	YIFDNVAK	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	485.26	1	y	4	431.26	3038.00	YIFDNVAK	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	489.26	1	y	7	814.45	5000.00	YIFDNVAK	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	489.26	1	y	6	701.37	377300.00	YIFDNVAK	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	489.26	1	y	5	554.30	28990.00	YIFDNVAK	SPIKE_IN
sp|Q15293|RCN1_HUMAN	YIFDNVAK	28.66	2	489.26	1	y	4	439.28	3038.00	YIFDNVAK	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	850.92	1	y	10	1151.58	894500.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	850.92	1	y	9	1052.52	4557000.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	850.92	1	y	8	905.45	2730000.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	850.92	1	y	7	834.41	4932000.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	855.92	1	y	10	1161.59	894500.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	855.92	1	y	9	1062.52	4557000.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	855.92	1	y	8	915.46	2730000.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	EGVYTVFAPTNEAFR	71.94	2	855.92	1	y	7	844.42	4932000.00	EGVYTVFAPTNEAFR	SPIKE_IN
sp|Q15582|BGH3_HUMAN	LTLLAPLNSVFK	109.84	2	658.40	1	y	10	1101.67	962500.00	LTLLAPLNSVFK	SPIKE_IN
sp|Q15582|BGH3_HUMAN	LTLLAPLNSVFK	109.84	2	658.40	1	y	9	988.58	1978000.00	LTLLAPLNSVFK	SPIKE_IN

Supplementary table 3 contd.

sp|Q15582|BGH3_HUMAN	LTLLAPLNSVFK	109.84	2	658.40	1	y	8	875.50	5871000.00	LTLLAPLNSVFK	SPIKE_IN
sp|Q15582|BGH3_HUMAN	LTLLAPLNSVFK	109.84	2	662.41	1	y	10	1109.68	962500.00	LTLLAPLNSVFK	SPIKE_IN
sp|Q15582|BGH3_HUMAN	LTLLAPLNSVFK	109.84	2	662.41	1	y	9	996.60	1978000.00	LTLLAPLNSVFK	SPIKE_IN
sp|Q15582|BGH3_HUMAN	LTLLAPLNSVFK	109.84	2	662.41	1	y	8	883.51	5871000.00	LTLLAPLNSVFK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	VALAGLLGFGLGK	109.43	2	608.38	1	y	9	861.52	7713.00	VALAGLLGFGLGK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	VALAGLLGFGLGK	109.43	2	608.38	1	y	7	691.41	3708.00	VALAGLLGFGLGK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	VALAGLLGFGLGK	109.43	2	608.38	1	y	6	578.33	3376.00	VALAGLLGFGLGK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	VALAGLLGFGLGK	109.43	2	612.38	1	y	9	869.53	7713.00	VALAGLLGFGLGK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	VALAGLLGFGLGK	109.43	2	612.38	1	y	7	699.43	3708.00	VALAGLLGFGLGK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	VALAGLLGFGLGK	109.43	2	612.38	1	y	6	586.34	3376.00	VALAGLLGFGLGK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	332.50	1	y	4	428.25	392900.00	DAHFPPPSK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	332.50	1	y	3	331.20	68930.00	DAHFPPPSK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	332.50	1	b	4	471.20	184200.00	DAHFPPPSK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	332.50	1	b	5	568.25	487700.00	DAHFPPPSK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	335.17	1	y	4	436.26	392900.00	DAHFPPPSK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	335.17	1	y	3	339.21	68930.00	DAHFPPPSK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	335.17	1	b	4	471.20	184200.00	DAHFPPPSK	SPIKE_IN
sp|Q56VL3|OCAD2_HUMAN	DAHFPPPSK	-16.28	3	335.17	1	b	5	568.25	487700.00	DAHFPPPSK	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GLYWVAPLNTDGR	71.93	2	731.38	1	y	9	942.50	71970.00	GLYWVAPLNTDGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GLYWVAPLNTDGR	71.93	2	731.38	1	y	8	843.43	244600.00	GLYWVAPLNTDGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GLYWVAPLNTDGR	71.93	2	736.38	1	y	9	952.51	71970.00	GLYWVAPLNTDGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GLYWVAPLNTDGR	71.93	2	736.38	1	y	8	853.44	244600.00	GLYWVAPLNTDGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GFSYLYGAWGR	71.9	2	638.81	1	y	7	822.43	33300.00	GFSYLYGAWGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GFSYLYGAWGR	71.9	2	638.81	1	y	6	709.34	98950.00	GFSYLYGAWGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GFSYLYGAWGR	71.9	2	638.81	1	y	5	546.28	76040.00	GFSYLYGAWGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GFSYLYGAWGR	71.9	2	643.81	1	y	7	832.43	33300.00	GFSYLYGAWGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GFSYLYGAWGR	71.9	2	643.81	1	y	6	719.35	98950.00	GFSYLYGAWGR	SPIKE_IN
sp|Q6UX06|OLFM4_HUMAN	GFSYLYGAWGR	71.9	2	643.81	1	y	5	556.29	76040.00	GFSYLYGAWGR	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	TLPIGQNFPIR	60.42	2	628.36	1	y	9	1041.58	5443000.00	TLPIGQNFPIR	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	TLPIGQNFPIR	60.42	2	628.36	1	y	8	944.53	3437000.00	TLPIGQNFPIR	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	TLPIGQNFPIR	60.42	2	628.36	1	y	7	831.45	6764000.00	TLPIGQNFPIR	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	TLPIGQNFPIR	60.42	2	633.37	1	y	9	1051.59	5443000.00	TLPIGQNFPIR	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	TLPIGQNFPIR	60.42	2	633.37	1	y	8	954.54	3437000.00	TLPIGQNFPIR	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	TLPIGQNFPIR	60.42	2	633.37	1	y	7	841.46	6764000.00	TLPIGQNFPIR	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	STHYQQYQPVVTLQK	26.92	3	607.32	1	y	7	784.49	389200.00	STHYQQYQPVVTLQK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	STHYQQYQPVVTLQK	26.92	3	607.32	1	y	5	588.37	90510.00	STHYQQYQPVVTLQK	SPIKE_IN

Supplementary table 3 contd.

sp|Q8WUJ3|K1199_HUMAN	STHYQQYQPVVTLQK	26.92	3	607.32	1	y	4	489.30	78260.00	STHYQQYQPVVTLQK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	STHYQQYQPVVTLQK	26.92	3	609.99	1	y	7	792.51	389200.00	STHYQQYQPVVTLQK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	STHYQQYQPVVTLQK	26.92	3	609.99	1	y	5	596.39	90510.00	STHYQQYQPVVTLQK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	STHYQQYQPVVTLQK	26.92	3	609.99	1	y	4	497.32	78260.00	STHYQQYQPVVTLQK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	IFQVVPIPVVK	80.03	2	619.90	1	y	7	751.51	5630.00	IFQVVPIPVVK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	IFQVVPIPVVK	80.03	2	619.90	1	y	6	652.44	10820.00	IFQVVPIPVVK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	IFQVVPIPVVK	80.03	2	619.90	1	y	4	442.30	18970.00	IFQVVPIPVVK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	IFQVVPIPVVK	80.03	2	623.90	1	y	7	759.52	5630.00	IFQVVPIPVVK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	IFQVVPIPVVK	80.03	2	623.90	1	y	6	660.45	10820.00	IFQVVPIPVVK	SPIKE_IN
sp|Q8WUJ3|K1199_HUMAN	IFQVVPIPVVK	80.03	2	623.90	1	y	4	450.32	18970.00	IFQVVPIPVVK	SPIKE_IN
sp|Q92688|AN32B_HUMAN	ELVLDNCK	4.86	2	495.75	1	y	6	748.37	1554000.00	ELVLDNC[+57]K	SPIKE_IN
sp|Q92688|AN32B_HUMAN	ELVLDNCK	4.86	2	495.75	1	y	5	649.30	2750000.00	ELVLDNC[+57]K	SPIKE_IN
sp|Q92688|AN32B_HUMAN	ELVLDNCK	4.86	2	495.75	1	y	4	536.21	1829000.00	ELVLDNC[+57]K	SPIKE_IN
sp|Q92688|AN32B_HUMAN	ELVLDNCK	4.86	2	499.76	1	y	6	756.38	1554000.00	ELVLDNC[+57]K	SPIKE_IN
sp|Q92688|AN32B_HUMAN	ELVLDNCK	4.86	2	499.76	1	y	5	657.31	2750000.00	ELVLDNC[+57]K	SPIKE_IN
sp|Q92688|AN32B_HUMAN	ELVLDNCK	4.86	2	499.76	1	y	4	544.23	1829000.00	ELVLDNC[+57]K	SPIKE_IN
sp|Q92688|AN32B_HUMAN	LPNLTHLNLSGNK	47.19	3	474.27	1	y	7	745.42	1126000.00	LPNLTHLNLSGNK	SPIKE_IN
sp|Q92688|AN32B_HUMAN	LPNLTHLNLSGNK	47.19	3	474.27	1	y	6	632.34	1174000.00	LPNLTHLNLSGNK	SPIKE_IN
sp|Q92688|AN32B_HUMAN	LPNLTHLNLSGNK	47.19	3	474.27	1	y	4	405.21	625000.00	LPNLTHLNLSGNK	SPIKE_IN
sp|Q92688|AN32B_HUMAN	LPNLTHLNLSGNK	47.19	3	476.94	1	y	7	753.43	1126000.00	LPNLTHLNLSGNK	SPIKE_IN
sp|Q92688|AN32B_HUMAN	LPNLTHLNLSGNK	47.19	3	476.94	1	y	6	640.35	1174000.00	LPNLTHLNLSGNK	SPIKE_IN
sp|Q92688|AN32B_HUMAN	LPNLTHLNLSGNK	47.19	3	476.94	1	y	4	413.22	625000.00	LPNLTHLNLSGNK	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	728.39	1	y	10	989.52	31220.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	728.39	1	y	9	861.46	218700.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	728.39	1	y	7	707.38	21780.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	728.39	1	y	5	553.31	22490.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	733.39	1	y	10	999.52	31220.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	733.39	1	y	9	871.47	218700.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	733.39	1	y	7	717.39	21780.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	LGAPQQPGPGPPPSR	5.05	2	733.39	1	y	5	563.32	22490.00	LGAPQQPGPGPPPSR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	EAEYFELPELVR	90.7	2	747.88	1	y	8	1002.56	307900.00	EAEYFELPELVR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	EAEYFELPELVR	90.7	2	747.88	1	y	7	855.49	436000.00	EAEYFELPELVR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	EAEYFELPELVR	90.7	2	747.88	1	y	6	726.45	306500.00	EAEYFELPELVR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	EAEYFELPELVR	90.7	2	752.88	1	y	8	1012.57	307900.00	EAEYFELPELVR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	EAEYFELPELVR	90.7	2	752.88	1	y	7	865.50	436000.00	EAEYFELPELVR	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	EAEYFELPELVR	90.7	2	752.88	1	y	6	736.46	306500.00	EAEYFELPELVR	SPIKE_IN

Supplementary table 3 contd.

sp|Q96CX2|KCD12_HUMAN	FNFLEQAFDK	82.1	2	629.81	1	y	8	997.50	71810.00	FNFLEQAFDK	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	FNFLEQAFDK	82.1	2	629.81	1	y	7	850.43	53120.00	FNFLEQAFDK	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	FNFLEQAFDK	82.1	2	629.81	1	y	6	737.35	60780.00	FNFLEQAFDK	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	FNFLEQAFDK	82.1	2	633.82	1	y	8	1005.51	71810.00	FNFLEQAFDK	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	FNFLEQAFDK	82.1	2	633.82	1	y	7	858.44	53120.00	FNFLEQAFDK	SPIKE_IN
sp|Q96CX2|KCD12_HUMAN	FNFLEQAFDK	82.1	2	633.82	1	y	6	745.36	60780.00	FNFLEQAFDK	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	SNCYGYFR	21.01	2	533.72	1	y	6	865.37	1812000.00	SNC[+57]YGYFR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	SNCYGYFR	21.01	2	533.72	1	y	5	705.34	1645000.00	SNC[+57]YGYFR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	SNCYGYFR	21.01	2	533.72	1	y	4	542.27	2266000.00	SNC[+57]YGYFR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	SNCYGYFR	21.01	2	538.73	1	y	6	875.37	1812000.00	SNC[+57]YGYFR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	SNCYGYFR	21.01	2	538.73	1	y	5	715.34	1645000.00	SNC[+57]YGYFR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	SNCYGYFR	21.01	2	538.73	1	y	4	552.28	2266000.00	SNC[+57]YGYFR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	EASTIAEYISGYQR	65.67	2	794.39	1	y	9	1086.52	2749000.00	EASTIAEYISGYQR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	EASTIAEYISGYQR	65.67	2	794.39	1	y	7	886.44	905700.00	EASTIAEYISGYQR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	EASTIAEYISGYQR	65.67	2	794.39	1	y	5	610.29	397900.00	EASTIAEYISGYQR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	EASTIAEYISGYQR	65.67	2	799.39	1	y	9	1096.53	2749000.00	EASTIAEYISGYQR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	EASTIAEYISGYQR	65.67	2	799.39	1	y	7	896.45	905700.00	EASTIAEYISGYQR	SPIKE_IN
sp|Q9BYZ8|REG4_HUMAN	EASTIAEYISGYQR	65.67	2	799.39	1	y	5	620.30	397900.00	EASTIAEYISGYQR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	LITINQQWKPIEELQNVQR	79.06	3	784.10	1	y	10	1225.65	272300.00	LITINQQWKPIEELQNVQR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	LITINQQWKPIEELQNVQR	79.06	3	784.10	1	y	7	886.47	202400.00	LITINQQWKPIEELQNVQR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	LITINQQWKPIEELQNVQR	79.06	3	787.44	1	y	10	1235.66	272300.00	LITINQQWKPIEELQNVQR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	LITINQQWKPIEELQNVQR	79.06	3	787.44	1	y	7	896.48	202400.00	LITINQQWKPIEELQNVQR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	TDVILDR	3.9	2	416.23	1	y	6	730.41	9723.00	TDVILDR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	TDVILDR	3.9	2	416.23	1	y	5	615.38	271500.00	TDVILDR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	TDVILDR	3.9	2	421.24	1	y	6	740.42	9723.00	TDVILDR	SPIKE_IN
sp|Q9C002|NMES1_HUMAN	TDVILDR	3.9	2	421.24	1	y	5	625.39	271500.00	TDVILDR	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	NVTGHYISPFHDIPLK	53.66	3	613.32	1	y	10	1166.66	297900.00	NVTGHYISPFHDIPLK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	NVTGHYISPFHDIPLK	53.66	3	613.32	1	y	9	1053.57	418600.00	NVTGHYISPFHDIPLK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	NVTGHYISPFHDIPLK	53.66	3	613.32	1	y	8	966.54	252600.00	NVTGHYISPFHDIPLK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	NVTGHYISPFHDIPLK	53.66	3	616.00	1	y	10	1174.67	297900.00	NVTGHYISPFHDIPLK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	NVTGHYISPFHDIPLK	53.66	3	616.00	1	y	9	1061.59	418600.00	NVTGHYISPFHDIPLK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	NVTGHYISPFHDIPLK	53.66	3	616.00	1	y	8	974.55	252600.00	NVTGHYISPFHDIPLK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	LIAINANDPEASK	21.8	2	678.36	1	y	11	1129.55	1482000.00	LIAINANDPEASK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	LIAINANDPEASK	21.8	2	678.36	1	y	9	945.43	2066000.00	LIAINANDPEASK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	LIAINANDPEASK	21.8	2	678.36	1	y	5	531.28	3690000.00	LIAINANDPEASK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	LIAINANDPEASK	21.8	2	682.37	1	y	11	1137.56	1482000.00	LIAINANDPEASK	SPIKE_IN

Supplementary table 3 contd.

sp|Q9H2U2|IPYR2_HUMAN	LIAINANDPEASK	21.8	2	682.37	1	y	9	953.44	2066000.00	LIAINANDPEASK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	LIAINANDPEASK	21.8	2	682.37	1	y	5	539.29	3690000.00	LIAINANDPEASK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	AFALEVIK	54.38	2	445.77	1	y	7	819.50	174.90	AFALEVIK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	AFALEVIK	54.38	2	445.77	1	y	6	672.43	1540.00	AFALEVIK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	AFALEVIK	54.38	2	445.77	1	y	5	601.39	435.20	AFALEVIK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	AFALEVIK	54.38	2	449.78	1	y	7	827.51	174.90	AFALEVIK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	AFALEVIK	54.38	2	449.78	1	y	6	680.44	1540.00	AFALEVIK	SPIKE_IN
sp|Q9H2U2|IPYR2_HUMAN	AFALEVIK	54.38	2	449.78	1	y	5	609.41	435.20	AFALEVIK	SPIKE_IN
sp|Q9UKY7|CDV3_HUMAN	SLDNFFAK	45.12	2	471.24	1	y	6	741.36	2035000.00	SLDNFFAK	SPIKE_IN
sp|Q9UKY7|CDV3_HUMAN	SLDNFFAK	45.12	2	471.24	1	y	5	626.33	139900.00	SLDNFFAK	SPIKE_IN
sp|Q9UKY7|CDV3_HUMAN	SLDNFFAK	45.12	2	471.24	1	y	4	512.29	53090.00	SLDNFFAK	SPIKE_IN
sp|Q9UKY7|CDV3_HUMAN	SLDNFFAK	45.12	2	475.25	1	y	6	749.37	2035000.00	SLDNFFAK	SPIKE_IN
sp|Q9UKY7|CDV3_HUMAN	SLDNFFAK	45.12	2	475.25	1	y	5	634.34	139900.00	SLDNFFAK	SPIKE_IN
sp|Q9UKY7|CDV3_HUMAN	SLDNFFAK	45.12	2	475.25	1	y	4	520.30	53090.00	SLDNFFAK	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	GALQNIIPASTGAAK	44.65	2	706.40	1	y	11	1042.59	971600.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	GALQNIIPASTGAAK	44.65	2	706.40	1	y	10	928.55	750800.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	GALQNIIPASTGAAK	44.65	2	706.40	1	y	9	815.46	5160000.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	GALQNIIPASTGAAK	44.65	2	710.41	1	y	11	1050.60	971600.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	GALQNIIPASTGAAK	44.65	2	710.41	1	y	10	936.56	750800.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	GALQNIIPASTGAAK	44.65	2	710.41	1	y	9	823.48	5160000.00	GALQNIIPASTGAAK	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	LISWYDNEFGYSNR	72.02	2	882.40	1	y	9	1101.46	177100.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	LISWYDNEFGYSNR	72.02	2	882.40	1	y	8	986.43	79510.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	LISWYDNEFGYSNR	72.02	2	882.40	1	y	6	743.35	152300.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	LISWYDNEFGYSNR	72.02	2	887.41	1	y	9	1111.47	177100.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	LISWYDNEFGYSNR	72.02	2	887.41	1	y	8	996.44	79510.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P04406|G3P_HUMAN_mtd2	LISWYDNEFGYSNR	72.02	2	887.41	1	y	6	753.36	152300.00	LISWYDNEFGYSNR	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	HLEINPDHSIIETLR	50.57	3	596.32	1	y	7	831.49	1195000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	HLEINPDHSIIETLR	50.57	3	596.32	1	y	6	744.46	531500.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	HLEINPDHSIIETLR	50.57	3	596.32	1	y	5	631.38	624000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	HLEINPDHSIIETLR	50.57	3	599.66	1	y	7	841.50	1195000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	HLEINPDHSIIETLR	50.57	3	599.66	1	y	6	754.47	531500.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	HLEINPDHSIIETLR	50.57	3	599.66	1	y	5	641.39	624000.00	HLEINPDHSIIETLR	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	DQVANSAFVER	9.64	2	618.30	1	y	9	992.52	461800.00	DQVANSAFVER	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	DQVANSAFVER	9.64	2	618.30	1	y	8	893.45	775100.00	DQVANSAFVER	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	DQVANSAFVER	9.64	2	623.31	1	y	9	1002.52	461800.00	DQVANSAFVER	SPIKE_IN
sp|P07900|HS90A_HUMAN_mtd2	DQVANSAFVER	9.64	2	623.31	1	y	8	903.46	775100.00	DQVANSAFVER	SPIKE_IN

Supplementary table 4. Peptides identified with a false discovery rate (FDR) below 1% and a Cscore greater than 10 (SpectroDive output file).

Supplementary table 5. MSstats Input File (Method 1)

Supplementary table 6. MSstats Input File (Method 2)

(Tables were not added due to file size. They would be made available in the electronic version of the thesis and also through the supplementary links to the manuscript).

6 Conclusions and future perspectives

In the first part of this study, we report novel findings from an untargeted quantitative proteomic study on a large cohort of precancerous colorectal tissues. Whereas most proteomic studies on colorectal tumors have targeted advanced stage lesions, we focused on the precancerous lesions. This involved a comprehensive exploration of the proteome of colorectal adenomas and adjacent normal mucosa samples using iTRAQ 8plex chemical labelling and shotgun mass spectrometry. We achieved an appreciable depth in our coverage of the proteome of colorectal tissues and we identified a large set of proteins whose expression was dysregulated in colorectal adenomas compared to samples of the normal mucosa. Furthermore, alterations in key enzymes of the polyol pathway were observed to be peculiar to the precancerous neoplasms. This finding poses further questions on the involvement of the polyol pathway in glucose metabolism, during tumorigenesis. Up-regulated SORD expression and activity in adenomas would enhance the production of fructose. Fructose is also several times more effective than glucose in promoting intracellular non-enzymatic glycation. Advanced glycation end products have been proven to contribute to vascular complications in diabetes and aging. Whether these fructose-driven metabolic events play a role in the development of adenomas is unclear, but an increase in the expression of KHK, the enzyme that catalyzes the transformation of fructose to fructose-1-P, downstream from the polyol pathway was clearly apparent in the few adenomas we studied. There is a dearth of information associating the polyol pathway and non-enzymatic glycation to tumorigenesis.

Our preliminary data on the abundance of polyol pathway metabolites (glucose and sorbitol) in colorectal tissues revealed interesting findings that deserve additional investigations. Adenomas exhibited dramatically reduced concentrations of glucose, the initial substrate in the polyol pathway. Sorbitol level was insignificantly decreased in adenomas and also fructose showed slightly lower levels in adenomas. Advanced cancer cells consume glucose at a much higher rate than normal cells, and much of their intracellular energy (ATP) is generated by aerobic glycolysis rather than through oxidative phosphorylation of glucose in the mitochondria (i.e. the Warburg effect). Increased aerobic glycolysis facilitates ATP generation via the conversion of pyruvate to lactate. Hence higher levels of intracellular lactate are often associated with cancer cells unlike normal cells. The relative concentration of lactate in the adenomas we tested was significantly greater than those found in matched samples of normal

mucosa, suggesting that the Warburg effect is already evident in precancerous colorectal lesions. Studies involving metabolic flux analysis to monitor the fate of isotopic tracers in *in vitro* and *in vivo* systems would provide further insight into the biological roles of this pathway in tumorigenesis. More specifically, tracking the fate of glucose breakdown with isotope-labelled metabolites is crucial to understand the connection between glucose metabolism and metabolites of the polyol pathway.

The verification of candidate biomarkers for the early diagnosis of colorectal carcinogenesis remains a major challenge due to the lack of techniques for the sensitive, reproducible quantification of multiple proteins over large cohorts of samples. The early detection of colorectal cancer, at the adenoma or early adenocarcninoma stage, remains the best way to curb the high incidence and hence mortality of the disease.

Most of the candidate protein markers verified in the second part of this work were previously reported in a quantitative (iTRAQ 8plex) shotgun discovery study on human colorectal precancerous tissues. We describe the high-throughput development and optimization of reproducible selected reaction monitoring assays for 40 human proteins using heavy isotope-labelled peptides. The sample preparation and analytical protocols were optimized to ensure reproducibility and analytical precision of measurements. Our integrated SRM workflow combined software tools for SRM assay development (Skyline), automated peak group identification (mProphet in SpectroDive) and protein significance analysis (MSstats). We applied these assays to detect 28 proteins in 72 human colorectal tissues (19 adenoma and paired normal mucosa; 17 adenocarcinoma and paired normal mucosa). Furthermore, we show that these SRM assays permit reproducible quantification of candidate biomarker proteins by profiling 25 putative biomarker targets over 78 colorectal samples, including six technical duplicates. We effected sample randomization and blocking strategies to achieve reproducible quantification with minimal biological and technical variability. In addition to this, differences in the expression of 25 proteins were validated in precancerous (adenomas) and cancerous (adenocarcinomas) lesions respectively, compared to the normal mucosa. We confirmed the significant (fold change > 1.2, adjusted P value < 0.02) over expression of ten proteins (SORD, SPB5, ANXA3, REG4, S10AB, NUCL, NGAL, LDHA, G6PD, AN32A) in precancerous as well as in cancerous neoplasms of the colorectum. These protein changes are unambiguously indicative of protein abundance variations that could be detectable phenotypes for early stage tumors. There is need for further clinical verification studies to monitor the abundance of these

candidate markers in the plasma of colorectal patients. This would define their suitability for use in non-invasive diagnostic tests for colorectal tumorigenesis. We have initiated plans to monitor a selected number of tumor markers in plasma samples from colorectal cancer patients. We aim to apply our established SRM assays to detect and quantify these proteins in human plasma.

7 Abbreviations

GI	Gastrointestinal tract
CRC	Colorectal Cancer
ACF	Aberrant Crypt Foci
APC	Adenomatous Polyposis Coli
TVA	Tubulovillous
FAP	Familial Adenomatous Polyposis
HNPCC	Hereditary Nonpolyposis Colon Cancer
TNM	Tumor, Node, Metastasis
BMI	Body Mass Index
DCBE	Double-Contrast Barium Enema
FOBT	Faecal Occult Blood Test
FIT	Faecal Immunochemical Test
MIN	Microsatellite Instability
CIN	Chromosomal Instability
MMR	Mismatch Repair
2DE	Two-dimensional Gel Electrophoresis
MS	Mass Spectrometry
iTRAQ	Isobaric Tag for Relative and Absolute Quantification
iCAT	Isotope-Coded Affinity Tags
QQQ	Triple quadrupole
RT	Retention time
SORD	Sorbitol dehydrogenase
ATP	Adenosine Triphosphate
m/z	Mass-to-charge Ratio
MS1	First MS phase

MS2	Second MS phase
SILAC	Stable Isotope Labelling by Amino Acids in Cell Culture
OGE	Off Gel Electrophoresis
SRM	Selected Reaction Monitoring
PRM	Parallel Reaction Monitoring
ELISA	Enzyme-linked Immunosorbent Assay
LC-MSMS	Liquid chromatography-tandem mass spectrometry
FDR	False Discovery Rate

8 Appendix

8.1 Appendix I.

Research

Sorbitol Dehydrogenase Overexpression and Other Aspects of Dysregulated Protein Expression in Human Precancerous Colorectal Neoplasms: A Quantitative Proteomics Study*[S]

Anuli Uzozie‡, Paolo Nanni§, Teresa Staiano¶, Jonas Grossmann§, Simon Barkow-Oesterreicher§, Jerry W. Shay‖, Amit Tiwari‡, Federico Buffoli¶, Endre Laczko§, and Giancarlo Marra‡

Colorectal adenomas are cancer precursor lesions of the large bowel. A multitude of genomic and epigenomic changes have been documented in these preinvasive lesions, but their impact on the protein effectors of biological function has not been comprehensively explored. Using shotgun quantitative MS, we exhaustively investigated the proteome of 30 colorectal adenomas and paired samples of normal mucosa. Total protein extracts were prepared from these tissues (prospectively collected during colonoscopy) and from normal (HCEC) and cancerous (SW480, SW620, Caco2, HT29, CX1) colon epithelial cell lines. Peptides were labeled with isobaric tags (iTRAQ 8-plex), separated via OFFGEL electrophoresis, and analyzed by means of LC-MS/MS. Nonredundant protein families (4325 in tissues, 2017 in cell lines) were identified and quantified. Principal component analysis of the results clearly distinguished adenomas from normal mucosal samples and cancer cell lines from HCEC cells. Two hundred and twelve proteins displayed significant adenoma-related expression changes (q-value < 0.02, mean fold change *versus* normal mucosa ±1.4), which correlated ($r = 0.74$) with similar changes previously identified by our group at the transcriptome level. Fifty-one (~25%) proteins displayed directionally similar expression changes in colorectal cancer cells (*versus* HCEC cells) and were therefore attributed to the epithelial component of adenomas. Although benign, adenomas already exhibited cancer-associated proteomic changes: 69 (91%) of the 76 protein up-regulations identified in these lesions have already been reported in cancers. One of the most striking changes involved sorbitol dehydrogenase, a key enzyme in the polyol pathway. Validation studies revealed dramatically increased sorbitol dehydrogenase concentrations and activity in adenomas and cancer cell lines, along with important changes in the expression of other enzymes in the same (AKR1B1) and related (KHK) pathways. Dysregulated polyol metabolism might represent a novel facet of metabolome remodeling associated with tumorigenesis. ***Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.035105, 1198–1218, 2014.***

Colorectal cancer ranks third among the world's high-incidence cancers and is a leading cause of cancer-related death among older adults (1, 2). In the United States alone, projections for 2013 include 102,480 new cases and 50,830 deaths (2). Cancerogenesis in the large bowel begins with the transformation of the epithelial cell lining of the gut. Molecular alterations, mainly involving the WNT signaling pathway, render these cells hyperproliferative, and they form benign adenomatous tumors. The neoplasms are initially noninvasive (3, 4), and the vast majority remain that way. But as genetic and epigenetic anomalies continue to accumulate, the tumor cells' capacity for invasion and destruction of surrounding tissues increases. At some point, this process drives certain adenomas into the realm of frank malignancy, transforming them into adenocarcinomas.

Early diagnosis of colorectal tumors has been greatly facilitated by screening methods based on fecal analysis or colonoscopy, but both approaches have limitations (5–9). Better understanding of the molecular mechanisms underlying large bowel tumorigenesis could improve our chances of detecting these lesions in the adenomatous or localized adenocarcinomatous stage, when the chances of successful treatment are greater. Promising results for the detection and validation of potential cancer biomarkers are emerging from proteomic studies of cancer development (10). Relative to

From the ‡Institute of Molecular Cancer Research, University of Zurich, Zurich 8057, Switzerland; §Functional Genomics Center of the University of Zurich and ETH, Zurich 8057, Switzerland; ¶Gastroenterology and Endoscopy Unit, Hospital of Cremona, Cremona 26100, Italy; ‖University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75211

Received October 15, 2013, and in revised form, February 7, 2014
Published, MCP Papers in Press, February 24, 2014, DOI 10.1074/mcp.M113.035105

Author contributions: P.N. and G.M. designed research; A.U., P.N., A.T., and E.L. performed research; J.W.S. contributed new reagents or analytic tools; J.G., S.B., and E.L. analyzed data; A.U. and G.M. wrote the paper; T.S. and F.B. performed tissue sample and clinical data collection.

older gel-electrophoresis-based approaches, shotgun proteomics methods, particularly those that include pre-MS OFFGEL electrophoretic peptide fractionation (11), enhance the sensitivity, robustness, and reproducibility of these studies (12) and expand the proteome coverage to include proteins that are less abundantly expressed (13–16). Furthermore, with the aid of isobaric-tag peptide-labeling strategies, MS can also be used for the relative quantification of protein expression levels within a series of multiple human tissue samples (12, 17–19).

Thus far, only a few MS-based proteomics studies have examined human colorectal adenomas (reviewed in Refs. 9 and 20). We therefore decided to explore the proteome of a relatively large series of these precancerous lesions (each with a paired sample of normal colon mucosa) using quantitative shotgun MS with the widely used iTRAQ[1] peptide labeling technique (21, 22) and OFFGEL fractionation. Adenoma-related protein expression variations specific to the epithelial compartments of these lesions were identified with a novel approach, which involved comparing the human tissue proteome with that of colon epithelial cell lines. The results of these studies revealed several protein expression changes previously documented only in advanced colorectal cancers. They also disclosed several novel changes with potentially important roles in early-stage large bowel tumorigenesis, including the marked up-regulation of a key enzyme in the polyol pathway.

EXPERIMENTAL PROCEDURES

Human Tissue Samples and Cell Lines—Human colorectal tissues were prospectively collected from patients undergoing colonoscopy in the Istituti Ospitalieri of Cremona, Italy. Approval was obtained from the local ethics committee, and tissues were used in accordance with the Declaration of Helsinki. Each donor provided written informed consent to sample collection, data analysis, and publication of the findings. Progressive numbers were assigned to each patient to protect human confidentiality. The series comprised 30 colorectal adenomas, each with a paired sample of normal mucosa from the same colon segment, >2 cm from the lesion. Tissues were collected endoscopically, promptly frozen in liquid nitrogen, and stored at −80 °C.

Five colorectal cancer cell lines (HT29, Caco2, CX1, SW480, and SW620) were obtained from the Zurich Cancer Network's Cell Line Repository. All had been recently purchased from the American Tissue Culture Collection (Teddington, UK) and were certified as mycoplasma infection free. We cultured Caco2 and CX1 cells in Dulbecco's modified Eagle's medium; HT29 cells in McCoy's medium; and SW480 and SW620 cells in RPMI 1640 medium supplemented with 10% fetal bovine serum, L-glutamine, and 1% penicillin-streptomycin (Sigma, St. Louis, MO). The recently established line of immortalized human colon epithelial cells (HCEC) was obtained from J. W. Shay and grown as described elsewhere (23).

Protein Extraction from Tissues and Cell Lines—For MS studies, frozen tissue samples were quickly weighed and homogenized on ice (1 min of grinding, 1 min on ice, 1 min of grinding) in a Wheaton glass borosilicate grinder containing a solution of 100 mM triethylammonium bicarbonate (Sigma, St Louis, MO), 1X Complete EDTA-free Protease Inhibitor Mixture (Roche, Mannheim, Germany), 1 M urea, 5 mM β-glycerophosphate disodium salt hydrate, 1 mM sodium orthovanadate, and 5 mM sodium fluoride (Sigma). The efficiency of cell lysis was microscopically confirmed. The homogenates were then sonicated with a Bioruptor (Diagenode, Denville, NJ) (high power, five 10-s/10-s on/off cycles) and centrifuged (16,000*g* for 5 min at 4 °C). The supernatant containing the proteins was collected and stored at −80 °C.

Cells (grown to >80% confluence in 15-cm^2 dishes) were washed in PBS, covered with 250 μl of the buffer used for tissue sample homogenization (see above), detached from the dish with a cell scraper, and homogenized (25 passages through a 25-gauge needle). The efficiency of cell lysis was microscopically confirmed. Sonication and centrifugation were repeated as described above, and the protein concentration was determined via Bradford assay. Prior to MS analysis, a 5-μg sample of each protein extract was subjected to one-dimensional gel electrophoresis on a 12% bisacrylamide gel to assess protein integrity and extraction protocol reproducibility. The entire proteomic workflow, from tissue/cell processing to statistical analysis, is summarized in Fig. 1 and described in detail in the next five subsections.

For sorbitol dehydrogenase (SORD) assays (see below), >80% confluent cells were washed in PBS and covered with a solution consisting of 100 mM triethanolamine (Sigma) and 1X Complete EDTA-free Protease Inhibitor Mixture (Roche). (A simple buffer was used to reduce the risk of introducing anti-enzymatic substances into our extract.) Cells were then scraped and homogenized with 25 passages through a 25-gauge needle. Tissue samples were weighed and homogenized in a Wheaton glass borosilicate grinder containing the buffer described above. After centrifugation (16,000 *g*, 4 °C, 5 min), the supernatant was aliquoted and stored at −80 °C. Protein concentration was measured via Bradford assay.

Protein Digestion and iTRAQ 8-plex Labeling—iTRAQ 8-plex experiments were performed to analyze tissue extracts (10 experiments) and cell-line extracts (1 experiment) (Fig. 1). Labeling efficiency and relative quantitation accuracy were assessed with the aid of two reference protein extract mixtures: one for tissue samples (pooled extracts from three normal tissues and three adenomas) and one for cell lines (pooled aliquots of each of the six cell line extracts). Fifty micrograms of protein per sample were used for each iTRAQ channel. Tryptic digestion (10% w/w, sequencing-grade modified trypsin, Promega, Madison, WI) and iTRAQ 8-plex labeling (AB Sciex, Framingham, MA) were performed according to the manufacturers' instructions (2.5-h incubation of samples with iTRAQ labels). For tissue experiments, two iTRAQ labels, 113 and 114, were chosen for the reference mixture, and labels 115/116, 117/118, and 119/121 were used for the three pairs of normal/adenomatous tissues included in each experiment. For the cell line experiment, labels 113 and 114 were used for the reference mixture, and labels 115–121 represented HCEC, HT29, Caco2, CX1, SW480, and SW620 cells, respectively (Fig. 1). After iTRAQ labeling, the samples (for each experiment) were combined, desalted on 500-mg SepPak C18 columns (Millipore, Billerica, MA), dried in a SpeedVac concentrator (Thermo Scientific), and subjected to peptide fractionation.

OFFGEL Electrophoresis—Peptide fractionation was performed according to the manufacturer's protocols with an Agilent 3100 OFFGEL fractionator and 12-well OFFGEL kit (both from Agilent Technologies, Santa Clara, CA). Briefly, samples were resolubilized in 1.8 ml of 1X OFFGEL peptide stock solution containing carrier ampholytes (pH range 3–10), loaded into the wells (150 μl per well), and focused until 20 kV/h was reached with a maximum current of 50 μA. For each experiment, 12 fractions were collected. A 15-μl aliquot of each fraction was acidified with 1.5 μl of a 50% acetonitrile/1%

[1] The abbreviations used are: iTRAQ, isobaric tags for relative and absolute quantification; HCEC, human colon epithelial cell; FDR, false discovery rate; GO, Gene Ontology; INHAT, inhibitor of acetyltransferases; MS/MS, tandem mass spectrometry; PSM, peptide spectra match; emPAI, exponentially modified protein abundance index; SORD, sorbitol dehydrogenase; TBST, Tris-buffered saline with Tween 20.

trifluoroacetic acid solution, desalted using ZipTip C18 (Millipore, Billerica, MA), dried, resolubilized in 15 μl of a 0.1% formic acid/3% acetonitrile solution, and analyzed with MS.

Liquid Chromatography and Mass Spectrometry—Peptide samples (4 μl) were analyzed on an LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to a nano-HPLC system (Eksigent Technologies, Dublin, CA). The solvent compositions were 0.2% formic acid and 1% acetonitrile for channel A and 0.2% formic acid and 80% acetonitrile for channel B. Peptides were loaded onto an in-house-made tip column (75 μm × 80 mm) packed with reverse-phase C18 material (AQ, 3 μm, 200 Å, Bischoff GmbH, Leonberg, Germany) and eluted (flow rate, 250 nL/min; solvent B gradient: from 3% to 30% in 62 min, from 30% to 45% in 70 min, and from 45% to 97% in 75 min).

Full-scan MS spectra (300–1700 *m/z*) were acquired at a resolution setting of 30,000 at 400 *m/z* after accumulation to a target value of 1×10^6. For the eight most intense signals per cycle above a threshold of 1000, both collision-induced dissociation and higher-energy collisional dissociation spectra were acquired in a data-dependent manner (Fig. 1). Collision-induced dissociation scans were recorded in the ion trap (settings: normalized collision energy, 35; maximum injection time, 50 ms; automatic gain control, 1×10^4 ions). For the higher-energy collisional dissociation scans, spectra were recorded at a resolution setting of 7500 at 400 *m/z* (normalized collision energy, 52; maximum injection time, 125 ms; automatic gain control, 5×10^4 ions). Charge state screening was enabled, and singly charged states were rejected. Precursor masses previously selected for MS/MS were excluded from further selection for 60 s, and the exclusion window was set at 10 ppm. The maximum number of entries in the exclusion list was set at 500. All samples were analyzed in duplicate, and precursors selected in the first run were excluded from fragmentation in the second run. The exclusion list was set on a time window of 4 min and a mass width of 10 ppm. Spectra were acquired using internal lock mass calibration on *m/z* 429.088735 and 445.120025.

Peak List Generation and Database Search—As depicted in Fig. 1, Mascot Distiller 2.4.3.3 (Matrix Science, Boston, MA) was used to generate Mascot generic format peak lists. De-isotoping and peak picking were not performed between 112.5 and 121.5 *m/z* (the range containing iTRAQ reporter ions), and the higher-energy collisional dissociation and collision-induced dissociation spectra were merged by summing. For each of the 11 experiments, the corresponding 24 Mascot generic format peak lists were concatenated and searched, with the aid of Mascot Server 2.3.02 (Matrix Science), against a forward UniProtKB/Swiss-Prot database for human proteins concatenated to a reversed decoyed FASTA database. The concatenated database contained a total of 147,438 proteins with accessions in Gene Ontology–compatible format and 260 common MS contaminants (NCBI taxonomy I.D. 9606, released December 13, 2011).

Methylthio (C), iTRAQ 8-plex labeling at the N terminus and lysine were set as fixed modifications, and variable modifications consisted of methionine oxidation and iTRAQ 8-plex labeling of tyrosine. We used the iTRAQ 8-plex-vs114 (Applied Biosystems Zug, Switzerland) quantitation method. The isotope and impurity correction factors used for each iTRAQ label were those provided by the manufacturer. Precursor and fragment tolerances were set at 10 ppm and 0.8 Da, respectively. The enzyme specificity was set to trypsin with an allowance of up to one missed cleavage. Using Mascot internal export scripts, we transformed Mascot DAT files into XML files and parsed them with in-house scripts so that peptide sequences, scores, and intensities of the individual reporter ion channels were reported. Confidently identified and quantified peptides were selected with the following filters: rank 1 (best spectra assignment); ion score, > 15; and presence of iTRAQ intensity values for reporter channels 113 and 114.

Quantification of Relative Protein Abundance—(These steps are described in the boxes of the lower half of Fig. 1.) Peptide reporter channel intensities were summed for each protein individually using R-scripts. Ratios were built from summed channels (113/114 to 121/114) for all proteins identified in each iTRAQ experiment. False discovery rates (FDRs) (24) were determined at the spectrum, peptide, and protein levels. The results of individual experiments were then merged into one matrix, which was used for statistical analysis in R and Perseus (Version 1.2.7.4). All proteins identified with the same peptide(s) were grouped into families, each of which was identified by a unique protein family number. Ratios of the intensity of each ion channel to that of 114 were converted to base 2 logarithmic values and normalized respectively on the median (which was set at 0), resulting in ratios that followed a Gaussian distribution. Proteins identified on the basis of the same peptide(s) shared the same family number and were represented once in our statistical analysis. The paired *t* test was used to compare the expression of a given protein in each adenoma and that found in the corresponding sample of normal mucosa. To correct for multiple comparisons, the FDR was controlled with the Benjamini–Hochberg procedure. The average protein-expression fold change in adenomas, relative to the normal mucosa, was then calculated. For this, median normalized ratios for all proteins in each paired adenoma–normal mucosa sample were deconvoluted of the reference standard effects (114) to compute the adenoma *versus* normal mucosa ratio per protein (deconvoluted fold change, (116/114)/(115/114) = (116/115)) and the mean fold change per protein in all tissue pairs. The Mascot emPAI values for all proteins were included in XML exports for each experiment. Thereafter, the mean Mascot emPAI value was calculated for all proteins.

Functional Annotation of Proteins—Gene Ontology (GO) annotations and GO terms for proteins in the UniProt/Swiss-Prot database were sourced from the European Bioinformatics Institute. The Scaffold program (Version 3) was used to identify the cellular localizations and biological processes most represented in lists of proteins quantified in tissues and cell lines. The topGO Bioconductor software package in R was used to identify and screen for GO biological process categories displaying enrichment for proteins that were differentially regulated in adenomas (*versus* normal mucosa) (25). First, we prepared a "universe" comprising all the proteins quantified in our study, each matched to GO terms and annotations. This served as the "background." The "foreground" consisted of the list of significantly dysregulated proteins. The most significant GO terms were scored with the Eliminating Genes (*elim*) method (25).

Western Blotting—Proteins were separated on a 10% SDS-polyacrylamide gel and transferred to a hydrophobic polyvinylidene difluoride membrane (GE Healthcare, Amersham Biosciences Hybond-P PVDF membrane, Pittsburgh, PA) according to standard protocols (26). After 1 h of blocking with 5% nonfat dry milk in TBS with 1% Tween 20 (milk–TBST), membranes were incubated overnight with the primary antibody (anti-SORD (HPA040260, Sigma); anti-aldose reductase, AKR1B1 (GTX113381, GeneTex, Irvine, CA); anti-ketohexokinase, KHK (GTX109591, GeneTex)) diluted 1:1000 in milk–TBST and washed once with milk–TBST (20 min) and twice with TBST (20 min). After 1 h of incubation in horseradish-peroxidase-conjugated secondary antibody (anti-rabbit IgG, GE Healthcare) diluted 1:5000 in milk–TBST, membranes were washed once with milk (20 min) and twice with TBST (20 min). Enhanced chemiluminescence (Amersham Biosciences, catalog no. RPN2106) was used to detect immunoreactive proteins.

Immunostaining of Cells and Tissues—HT29 and HCEC cells were seeded (3×10^5 per well) on 22-mm × 22-mm cover slips in six-well plates and grown under standard conditions until cells reached 70% to 80% confluence. Cells were then washed twice with PBS and fixed in ethanol:methanol solution (50:50) for 10 min at room temperature. Fixed cells were permeabilized (10 min with 0.25% Triton X-100),

blocked (30 min in 10% goat serum (X0907, Dako, Glostrup, Denmark)), and incubated with primary antibody (rabbit polyclonal anti-SORD, HPA040260, Sigma, 1:100) for 18 h at 4 °C. After three washes with PBS, the cells were incubated for 1 h with secondary antibody conjugated to polymer-HRP anti-rabbit (Dako, EnVision+ System-HRP, catalog no. K4010). They were then washed three times in PBS and incubated for 15 min in the substrate-chromogen 3,3-diaminobenzidine tetrahydrochloride (Dako, EnVision+ System-HRP, catalog no. K4010). Cells were washed quickly with PBS and mounted on slides (EUKITT, O. Kindler, GmbH, Freiburg, Germany) for light microscopy (Leica Microsystems GmbH, Wetzlar, Germany). Images were examined and recorded with Leica Application Suite (V3.3.0) software.

SORD immunohistochemistry was performed as previously described (27). Tissue sections (normal colon and ileum, colorectal adenomas, and adenocarcinomas) were incubated for 24 h at 4 °C with primary antibody (rabbit polyclonal anti-SORD, HPA040261, Sigma) at a 1:100 dilution.

Measurement of SORD Activity—Total protein was extracted from cell lines and tissues as described above. SORD catalyzes the reversible conversion of D-sorbitol to D-fructose, with β-NADH as a cofactor.

$$\text{D-fructose} + \beta\text{-NADH} \xleftrightarrow{\text{SORD}} \text{D-sorbitol} + \beta\text{-NAD} \quad \text{(Eq. 1)}$$

SORD activity was quantified via continuous spectrophotometric rate measurement of the β-NAD formation rate (temperature 25 °C, pH 7.6, A_{340}, light path of 1 cm) in a Cary 50 Scan UV-visible spectrophotometer using the Cary Kinetics Application (both from Varian Inc., Palo Alto, CA) (28). The final reagent concentrations in a 1-ml cuvette were as follows: 78.33 mM triethonolamine, 183 mM D-fructose, 0.21 mM β-NADH, 0.033% (w/v) BSA. The absorbance reading was recorded when the enzyme was added. One unit of enzyme activity was defined as the amount of enzyme required per minute to convert 1.0 μM D-fructose to D-sorbitol at pH 7.6 at 25 °C. A mixture of reagents plus recombinant SORD was used as the positive control; negative controls consisted of the same reagent mixture with no recombinant SORD, with recombinant SORD but no D-fructose, or with recombinant SORD but no β-NADH.

Extraction and Quantification of Intracellular Metabolites by Targeted Gas Chromatography Coupled with MS—Frozen tissue (50 to 100 mg) was homogenized in 250 μl of ice-cold 80% methanol using a glass borosilicate grinder from Wheaton (Rockdale, UK). The homogenate was microscopically examined to ensure that it was cell free, and then it was transferred to Eppendorf vials and left on ice for 15 min to ensure efficient protein precipitation. After centrifugation (15,000*g* for 3 min at 0 °C), the supernatant was snap-frozen and stored at −80 °C, and the protein content of the pellet was determined via the Bradford method.

For gas chromatography coupled with MS (GC-TOF-MS), 10 μl of supernatant was transferred to a 1.5-ml Eppendorf tube, and internal standards ($^{13}C_1$-sorbitol, $^{13}C_1$-fructose, and $^{13}C_1$-glucose; 1.2 pM of each) were added. The samples were then dried overnight in a vacuum centrifuge (Concentrator 5301, Eppendorf AG, Wesseling-Berzdorf, Germany). Methoxyamine hydrochloride and N-methyl-N-(trimethylsilyl)trifluoroecetamide were used as derivatization reagents (29).

The derivatized metabolites and internal standards were subjected to GC-TOF-MS (GC 7890A, Agilent Technologies, Santa Clara, CA; GCT Premier Micromass, Waters, Manchester, UK) with an Rxi-5Sil MS Integra-Guard column (length, 30 m; internal diameter, 0.25 mm) and a film thickness of 0.25 μm (Restek, Bellafonte, PA). One microliter of each derivatized sample was injected in splitless mode on a baffled glass liner and transferred to the capillary column by rapid heating of the liner from 50 °C to 250 °C at a rate of 12 °C/s. For the separation of the metabolites, helium was used at a flow rate of 1 ml/min, and after an initial hold time of 2 min, a temperature gradient from 80 °C to 320 °C (rate = 8 °C/min) was applied. The TOF-MS was set to acquire centroided standard electron ionization mass spectra over a range of 50 to 600 *m/z* at a rate of three spectra per second. The GC-MS transfer line was heated to 280 °C. Dynamic range enhancement was activated. C_6ClF_5 was used as lock mass compound.

The MassLynx and QuanLynx programs (Waters, UK) were used to review and analyze the acquired data. The absolute concentrations of D-sorbitol, α and β D-fructose, and α and β D-glucose were calculated on the basis of the ratio of the intensity of specific fragments originating from the unlabeled compound to that of the added labeled analog (internal standard). These concentrations were used to estimate intracellular levels per milligram of tissue (adenoma *versus* normal mucosa). The relative concentration of lactate was estimated from the ratio of the intensity of specific fragments originating from the unlabeled compound to that of the added $^{13}C_1$-sorbitol (internal standard).

RESULTS

Proteomic Analysis of Human Colorectal Tissues and Colon Cell Lines—We used a quantitative-MS-based discovery strategy to explore the proteome of human colorectal tissues and colon cell lines (normal and neoplastic). The characteristics of the precancerous colorectal lesions are listed in Table I. Protein extracts from these tumors and their paired samples of normal mucosa (60 samples total) were analyzed using iTRAQ LC-MS/MS and the workflow described in Fig. 1. The inclusion of two reference sample mixes allowed us to control for technical variability across the 10 experiments on tissue samples, as the reference sample was analyzed twice in each experiment. OFFGEL electrophoresis was used to obtain highly reproducible, pI-based, in-solution separation of pooled iTRAQ-labeled peptides. Furthermore, for relative quantification of proteins using iTRAQ reporter ions, we adopted a stringent FDR for peptide spectra matches (PSMs), and high-confidence peptides for protein quantification were selected only if the reporter ions (113 and 114) were quantified in the reference sample mix (iTRAQ reporter channels 113 and 114). The dataset generated with this approach was large and complex, but we developed a simplified analytical method that allowed us to work with and merge the large data files generated after MS/MS (Fig. 1). High-resolution MS/MS spectra acquired on the LTQ-Orbitrap Velos spectrometer after duplicate analysis of OFFGEL tissue sample fractions produced a total of 240 raw files (10 experiments, 120 fractions, 2 replicates). A total of 37,184 (FDR = 0.9%) unique tryptic peptides were confidently identified and quantified from 285,929 unique PSMs (FDR = 0.2%) (Table II, supplemental Table S1). Ten thousand four hundred and fifty-two proteins (FDR = 1.5%) were assembled from the quantified peptides. Proteins that were indistinguishable in MS/MS (*i.e.* two or more proteins identified on the basis of the same peptide sequence; see "Experimental Procedures" for details) were represented as a single family. The result was a total of 4325

TABLE I
Characteristics of the precancerous colorectal lesions included in the study

Patient number	Age	Sex	Colon segment involved	Maximum lesion diameter (mm)	Paris classification[a]	Pit pattern classification[b]	Microscopic appearance	Highest degree of dysplasia in the lesion[c]	Number of lesions at study colonoscopy[d]	Number of previously excised lesions
1	77	M	S	25	IIa+IIc	IIIs-IIIL	TA	LGD	1	0
2	73	F	A	25	IIa+IIc	IIIs-IIIL	TA	LGD	1	2
3	59	M	T	30	IIa+IIc	IIIs+IIIL	TA	LGD	1	0
4	73	F	R	50	Is	IV	VA	LGD	1	0
5	74	M	R	40	Is	IV	VA	HGD	2	1
6	77	M	C	25	IIa	IIIL	VA	LGD	1	0
7	80	M	A	40	IIa	IIIL	TVA	LGD	1	1
8	82	M	A	15	IIa	IIIL	VA	LGD	2	0
9	73	F	S	20	Ip	IV	TVA	LGD	1	0
10	70	F	C	25	IIa	IIIL	TVA	LGD	2	0
11	63	M	A	45	Is	IIIL-IV	TVA	LGD	0	0
12	68	M	A	30	IIa+Is	IIIL-IV	TVA	HGD	0	0
13	60	M	D	30	Is	IV-Vi	TVA	HGD	0	0
14	55	M	C	25	IIa+Is	IIIL-IV	SSA	LGD	0	0
15	70	M	A	15	Is	IV	TVA	LGD	7	0
16	85	F	S	25	Is+IIa	IV	TA	LGD	1	1
17	66	M	A	30	IIa	IIIL	TA	HGD	2	0
18	72	M	A	30	Is	IV	TVA	HGD	2	0
19	71	M	S	30	IIa	IIIL	TVA	LGD	2	0
20	59	M	R	60	Is	IV-Vi	TVA	HGD	1	0
21	78	M	A	50	Is	IV-Vi	TA	LGD	1	0
22	75	M	R	25	Is	IV-Vn	TVA	HGD	6	0
23	73	F	D	25	Is	IV	TA	LGD	1	0
24	69	F	R	90	Is+IIa	IV	TVA	LGD	1	0
25	75	M	T	18	IIa	IIIL	TA	LGD	1	0
26	61	M	A	40	Is+IIa	IV	TVA	LGD	20	0
27	76	M	S	30	Is	IV-Vi	TA	HGD	1	0
28	78	F	R	60	IIa+Is	IV	TVA	LGD	1	1
29	89	M	R	30	Is	IV	TA	LGD	3	0
30	75	M	A	50	Is	IV-Vn	TVA	HGD/cancer	7	0

Abbreviations: M, male; F, female; C, cecum; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum; TA, tubular adenoma; TVA, tubulovillous adenoma; VA, villous adenoma; SSA, sessile serrated adenoma; LGD, low-grade dysplasia; HGD, high-grade dysplasia.

[a] Macroscopic appearance of neoplastic lesions was classified according to the Paris Endoscopic Classification of Superficial Neoplastic Lesions (*Gastrointest. Endosc.* 2003;58(Suppl.):S3–S27).

[b] Morphological analysis of colon crypt patterns according to the Kudo classification (82).

[c] Low-grade versus high-grade dysplasia as defined by the World Health Organization classification of tumors of the digestive system at the editorial and consensus conference in Lyon, France, November 6–9, 1999.

[d] This number includes the lesion included in our proteomic study.

nonredundant protein families, two-thirds (2865, 66%) of which were relatively quantified in at least 9 normal mucosa–adenoma pairs, and 1072 (25%) in all 30 pairs (Table II, supplemental Table S1).

To verify the efficiency of iTRAQ protein labeling, we repeated the database search with methylthio (C) set as a fixed modification and iTRAQ 8-plex (N-term), iTRAQ 8-plex (K), iTRAQ 8-plex (Y), and oxidation (M) set as variable modifications. (All other search parameters were unchanged.) The assigned PSMs were filtered, as described in "Experimental Procedures," and the average iTRAQ labeling efficiency achieved in each of the 10 tissue experiments was 96% (supplemental Table S2). To ascertain the efficacy of including a standard sample mix as a reference for normalization, we compared combined Gaussian plots of $\log_2$-protein ratios of normal mucosa or adenoma samples with the respective reference channel per experiment (*e.g.* 115/114 *versus* 113/114 for normal tissues, 116/114 *versus* 113/114 for adenomas; see Fig. 1). The ratios displayed normal distributions in all channels. For the reference channel (113/114), $\log_2$-ratios were largely centered on 0, whereas the distribution of adenoma and normal channel $\log_2$-ratios was broader and not always centered at 0 (data not shown).

Sample complexity is a common problem in the analysis of proteomic data from human colorectal tissues. It stems in part from contamination of the epithelial cell proteome by proteins from stromal cells (which were inevitably present in our specimens, even though the endoscopic tissue sampling procedure we used yielded superficial specimens with consistently high epithelial contents). Microdissection can be utilized to

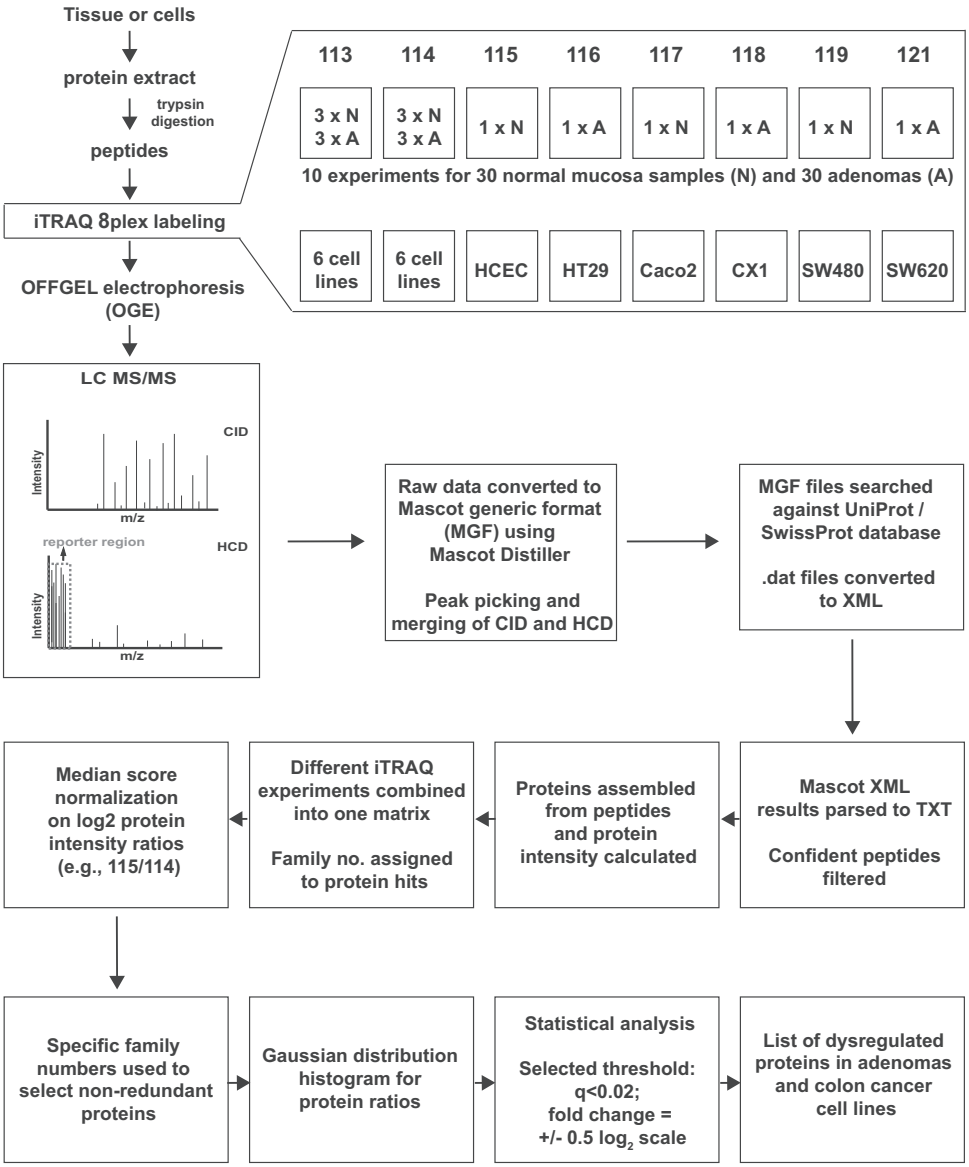


FIG. 1. **Project design and iTRAQ 8-plex labeling scheme.** Sample preparation for shotgun MS/MS and important steps in the analysis of proteomic data for the detection of dysregulated proteins in adenomas and colon cancer cell lines. For each experiment on tissue samples, iTRAQ tags were assigned to a duplicate reference (two identical pools of normal and adenoma samples: 113 and 114, respectively), normal tissues (115, 117, 119) and corresponding adenomas (116, 118, 121). The same pattern was repeated in all 10 experiments. In cell line experiments, two identical pools, each comprising all six cell lines, were used as references (113 and 114), and each of the remaining six tags was used to label a single cell line. The data analysis flow chart depicted in this figure is described in "Experimental Procedures."

isolate subpopulations of cells, but it can diminish the quantity and quality of the proteins, rendering them suboptimal for some types of proteomic analysis. To avoid this problem, we adopted a novel strategy for preliminary identification of the proteomic alterations that were *most likely* to involve the epithelial-cell component of the adenomas. The proteomic profiles of the colon tissues were compared with those of six colon epithelial cell lines (five colon cancer cell lines plus

TABLE II
Summary of proteomics data

	Tissues (n = 60)		Cell lines (n = 6)	
	Total	FDR (%)	Total	FDR (%)
Peptide spectra matches	285,929	0.2	27,922	0.4
Peptides	37,184	0.9	11,266	0.5
Proteins[a]	10,452	1.5	5056	1.1
Proteins[b]	4325	-	2017	-
Proteins[c]	1072	-	1957	-

[a] Total number of proteins quantified in the 10 tissue experiments and the single experiment with cell lines.
[b] Nonredundant protein families quantified in our dataset.
[c] Nonredundant protein families quantified in all 60 tissues or in all six cell lines.

HCEC cells, to our knowledge the only well-characterized line established from normal colorectal epithelium (23)). Changes in expression levels observed in adenomas (*i.e.* up-regulation or down-regulation with respect to normal mucosal levels) were presumed to be epithelial-cell-specific if similar changes were found in the colon cancer cell lines (relative to HCEC cells). After OFFGEL fractionation, duplicate MS analysis of iTRAQ-labeled peptides (24 fractions) from the six cell lines was performed in an LTQ-Orbitrap Velos mass spectrometer, and 11,266 peptides (FDR = 0.5%) were confidently identified and quantified from 27,922 unique PSMs (FDR = 0.4%) (Table II, supplemental Table S3). A total of 2017 nonredundant protein families (FDR = 1.1%) were identified and relatively quantified in cell lines; 1957 (97%) were present in all six cell lines (Table II). In the iTRAQ experiment with cell lines, the peptide labeling efficiency was 95% (supplemental Table S2).

Relative Quantification of the Proteomes of Colorectal Tissues and Cell Lines—The concentration range for proteins expressed in human tissues spans 10 orders of magnitude. We chose not to deplete our protein samples of high-abundance proteins (*e.g.* albumin, IgG), because with the number of tissue samples being analyzed, additional sample preparation steps were considered potential sources of confounding variability (30). As an alternative, each of the 10 pooled iTRAQ-labeled samples (10 experiments) was separated into 12 fractions based on the isoelectric point of peptides, reducing the complexity of our protein matrix and limiting the risk of bias toward the more abundant proteins.

The expression levels of the 4325 nonredundant protein families we were able to relatively quantify in colorectal tissues spanned 4 orders of magnitude, as deduced from the protein Mascot emPAI value (used as a proxy for the emPAI value (31) to estimate protein concentrations) (Fig. 2*A*). Thirty percent (1304/4325) of these families were relatively quantified on the basis of more than one unique peptide. At the top of this list were the large proteins AHNAK, DYNC1H1, DSP, and FCGBP (Fig. 2*B*). In colon epithelial cell lines, 1174 of the 2017 protein families were relatively quantified with more than one unique peptide.

Gene Ontology Annotation in Scaffold was used to identify the subcellular localizations of these protein families and the biological processes they were involved in. The GO categories represented in the tissue and cell line proteomes were fairly similar. In the cell line proteome, however, the categories generally contained fewer proteins, as the total number of proteins detected in these cells was less than that in the tissues (Fig. 2*C*). Cytoplasmic and organelle- or membrane-associated proteins were the most highly represented categories in our extracts, but nuclear proteins were also readily identified, which indicates that our protein extraction procedure was not strongly biased toward a few cell compartments. The most highly represented biological processes in the tissue proteome were metabolic or biosynthetic processes, whereas cell component organization and developmental processes predominated in the cell line proteome (Fig. 2*D*). Stromal contamination was probably responsible for the increased representation of immune system processes in the tissue proteome (relative to that of the cell lines).

Log_2-expression levels of the protein families identified in all tissues (n = 1072) and cell lines (n = 1957) (Table II) were subjected to principal component analysis, which easily distinguished the adenomas from the normal mucosa samples (Fig. 3*A*) and the five colon cancer cell lines from the immortalized normal colon epithelial cell line HCEC (Fig. 3*B*). The cancer cell lines were also segregated into three distinct groups reflecting their patient origins (Fig. 3*B*). When principal component analysis was performed on the expression intensity values of the 1496 nonredundant proteins expressed and quantified in all tissues and cell lines (*i.e.* those representing the intersection of the tissue (n = 10,452) and cell line (n = 5056) protein sets reported in Table II), colon cancer cell lines clustered with adenomas, whereas HCEC cells were closer to the normal mucosa samples (Fig. 3*C*).

As a quality control measure, data for the 60 tissue samples (1072 protein families) were subjected to hierarchical clustering analysis. As shown in supplemental Fig. S1, three main clusters emerged: one consisting almost exclusively of normal mucosa samples, a second containing mainly adenomas, and a third that included both tissue types. The 18 samples in the third cluster (nine adenoma–normal mucosa pairs) formed three subclusters, which corresponded to 3 of the 10 experiments for which trypsin digestion, iTRAQ labeling, and LC-MS/MS were performed on the same day. These findings were suggestive of an experimental bias. Indeed, when these 18 potentially substandard samples were included in subsequent statistical analyses, they diminished the stringency of our threshold and increased the error margin for false identification. We therefore excluded these samples from the analyses described in the following section.

Proteins Displaying Dysregulated Expression in Colorectal Adenomas and Colon Cell Lines—To identify proteins with significantly altered expression in adenomas (relative to normal mucosa), we analyzed data on the proteins quantified in

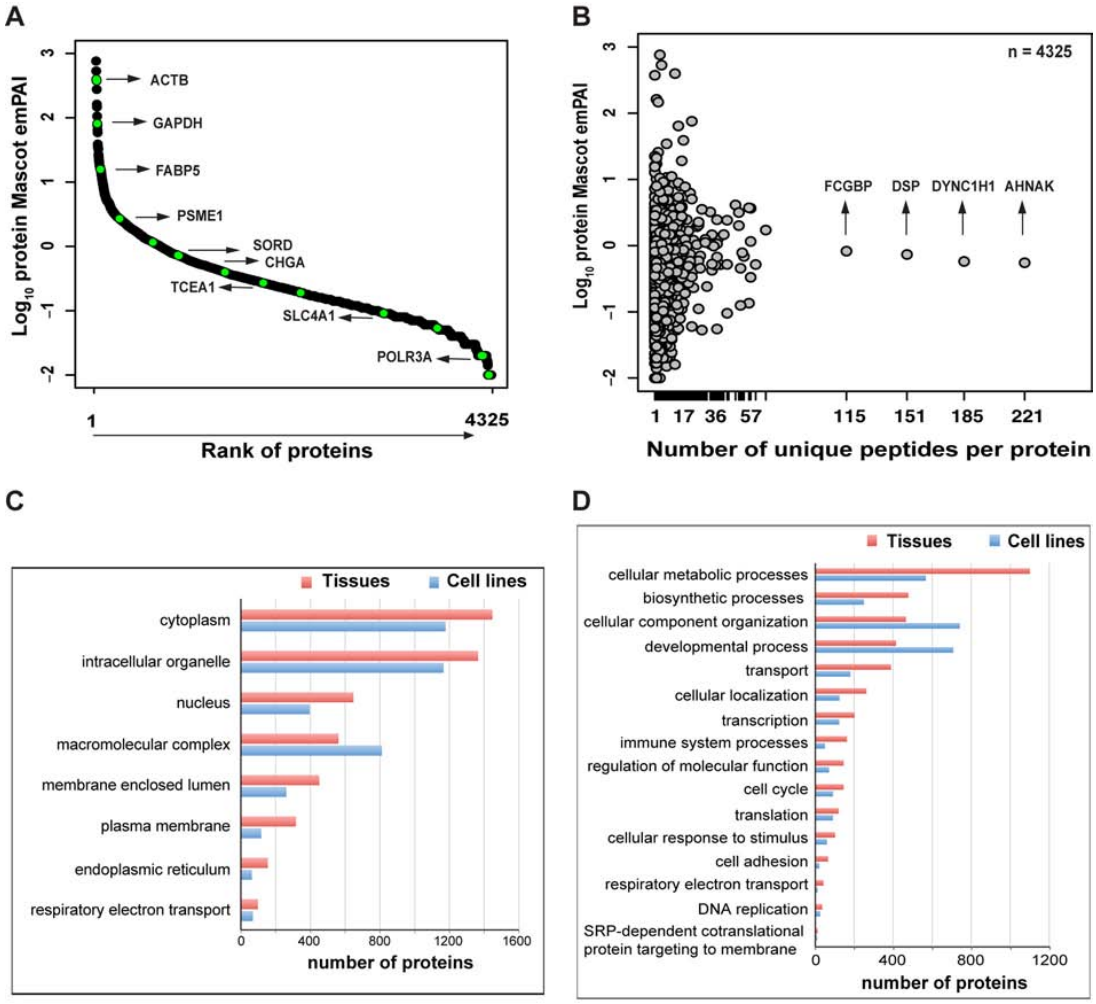


FIG. 2. **Protein coverage with iTRAQ shotgun analysis in colorectal tissues.** *A*, analysis of Mascot emPAI values (used as a proxy for emPAI values) revealed a dynamic range of protein abundance in tissues that spanned 4 orders of magnitude (*y*-axis) and corresponded with known abundance estimates for various proteins in these tissues. The high/moderate-abundance proteins (*e.g.* ACTB, FABP5, CHGA) and low-abundance protein (*e.g.* POLR3A) relatively quantified in our samples are highlighted relative to their mean Mascot emPAI value. *B*, distribution of reported abundance ranges for the proteins with at least one unique peptide identified in tissues, and the high-molecular-weight proteins with the greatest number of unique peptides identified. Subcellular localizations of the proteins identified in colorectal tissues and cell lines (*C*) and biological processes in which these proteins are involved (*D*). This analysis was performed using Scaffold and Gene Ontology annotations (see "Experimental Procedures").

the remaining 21 tissue pairs. The experimentally derived protein fold-change threshold defining differential expression was based on comparison of the distributions of average intensity $\log_2$ ratios in the reference standard (113 *versus* 114, seven experiments) and in patient samples (adenoma *versus* normal, seven experiments). The average ratios in the reference sample were centered on 1 (*i.e.* $\log_2$ 0). Average fold-change ratios for the tissue samples displayed wider variance (supplemental Fig. S2). Seventeen percent of the $\log_2$ ratios for the tissue samples exceeded $\pm$0.5 $\log_2$ scale (indicating a linear fold change $\geq \pm$1.4), as opposed to only 5% of those for the reference samples. For each protein, a paired *t* test was used to compare the intensity ratios in normal and adenomatous samples (*i.e.* normal/114; adenoma/114). After adjustment for multiple comparison (Benjamini–Hochberg method), we selected a stringent q value cutoff of $\leq$0.02.

The 212 proteins that satisfied this criterion and presented a mean expression fold change of $\pm$1.4 ($\log_2$ 0.5) or greater were classified as significantly dysregulated in adenomas. They included 76 with up-regulated expression and 136 with down-regulated expression in the tumor samples (Table III). When protein abundance iTRAQ ratios for these 212 proteins

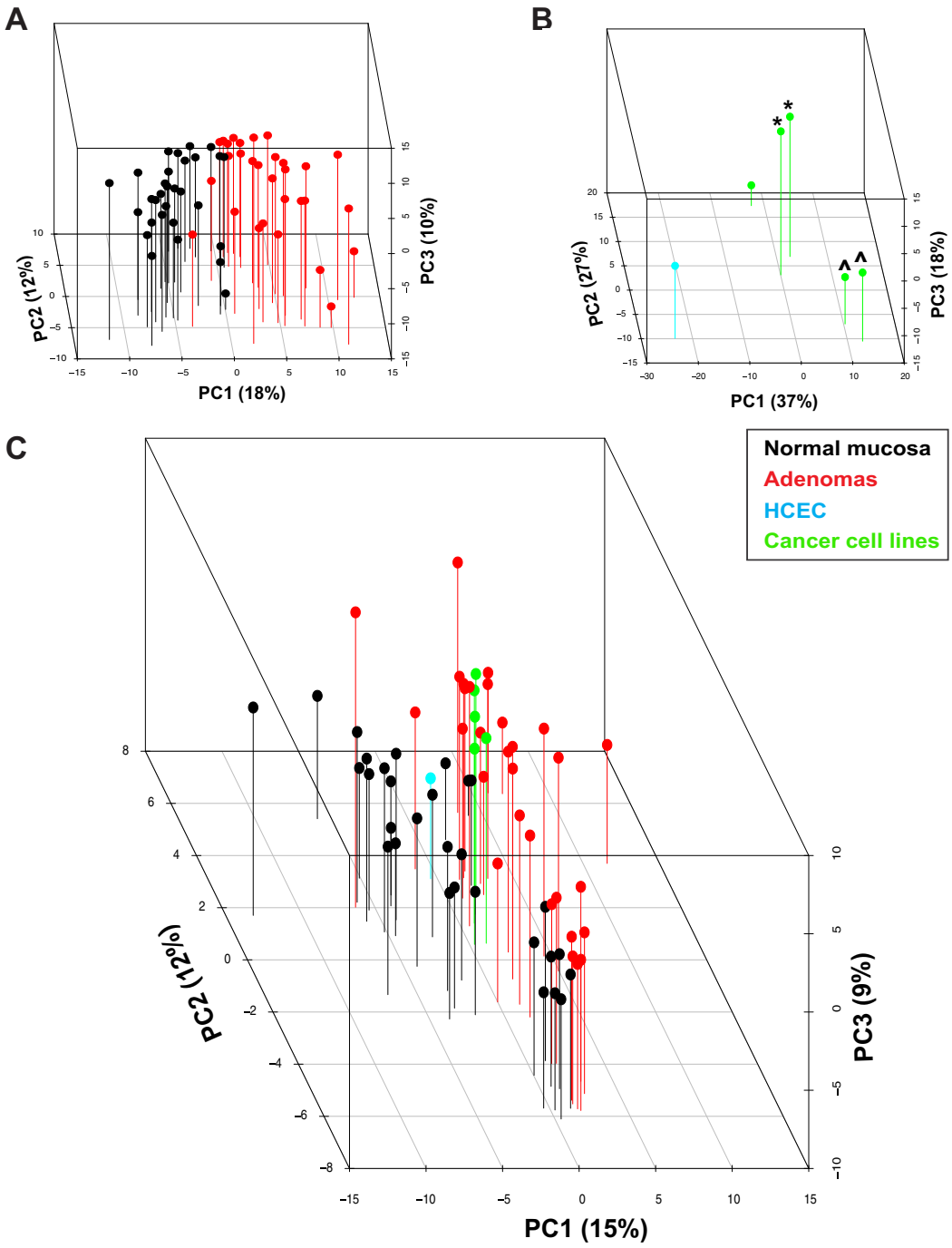


FIG. 3. **Principal component analysis of protein expression.** Three-dimensional principal component analysis score plot of log2 protein expression intensity values for (*A*) tissues (normal mucosa, black; adenomas, red), (*B*) cell lines (HCEC, cyan; colon cancer cell lines, green), and (*C*) both. The first three principal components (PCs) account for 40%, 82%, and 36% of the total variance in the tissue, cell line, and tissue + cell line sets, respectively. PC1, the main direction of spread in the three groupings, reflects intergroup variance based on tissue or cell line type (*i.e.* normal/immortalized *versus* tumorous). Cell lines derived from the same patient: *SW480 and SW620 cells; ^HT29 and CX1 cells.

TABLE III
Proteins displaying differential expression in adenomas versus normal mucosa

UniProt accession number[a]	Gene name	q value	Average fold change (log2)
P12429	ANXA3	0.00000001	1.44
Q9UN36	NDRG2	0.00000001	−0.79
P00918	CA2	0.00000003	−2.26
P23946	CMA1	0.00000005	−1.38
P00488	F13A1	0.00000005	−1.30
Q9UBR2	CTSZ	0.00000016	−1.23
P10645	CHGA	0.00000016	−1.82
O60844	ZG16	0.00000024	−2.02
P17174	GOT1	0.00000045	−0.61
P31949	S100A11	0.00000062	1.49
P00338[b]; **P07195**	LDHA; LDHB	0.00000068	0.62
O60701	UGDH	0.00000104	−0.70
P55011	SLC12A2	0.00000115	0.85
O95571	ETHE1	0.00000120	−0.74
Q01105	SET	0.00000230	0.72
Q16851	UGP2	0.00000275	−0.53
Q00796	SORD	0.00000275	0.62
P20231; Q15661	TPSB2; TPSAB1	0.00000275	−1.51
Q15181; **Q9H2U2**[b]	PPA1; PPA2	0.00000421	0.64
Q9H3G5	CPVL	0.00000467	−0.69
P01282	VIP	0.00000722	−1.10
P07339	CTSD	0.00000856	−0.71
P19338	NCL	0.00000934	0.50
Q6UWP2	DHRS11	0.00001070	−1.20
P04066	FUCA1	0.00001240	−1.32
Q53EL6	PDCD4	0.00001270	−0.51
P07585	DCN	0.00001580	−1.36
P02511	CRYAB	0.00002720	−1.19
Q96CX2	KCTD12	0.00003590	−0.70
Q05707; **P08123**	COL14A1	0.00004020	−1.21
P51884	LUM	0.00004160	−0.96
Q15063	POSTN	0.00004240	−2.10
P21397	MAOA	0.00004240	−0.63
O00748	CES2	0.00004610	−0.88
Q56VL3	OCIAD2	0.00004960	0.95
Q9BYZ8	REG4	0.00004960	0.98
P55008	AIF1	0.00005650	−0.75
P50224; **P50225**	SULT1A3; SULT1A1	0.00005760	−0.63
O14773	TPP1	0.00006430	−0.52
Q16853	AOC3	0.00006650	−1.16
P53634	CTSC	0.00006940	−0.58
O95881	TXNDC12	0.00006940	0.55
O75795	UGT2B17	0.00006940	−1.57
O00391	QSOX1	0.00006940	−0.70
Q99538	LGMN	0.00006940	−0.63
P12111	COL6A3	0.00006990	−0.74
P80188	LCN2	0.00006990	1.32
P12956[b]	XRCC6	0.00007190	0.53
Q6NZI2	PTRF	0.00007860	−0.76
P09382	LGALS1	0.00007880	−0.84
P25815	S100P	0.00008020	3.38
Q15118	PDK1	0.00008120	−0.90
O75380	NDUFS6	0.00008650	−0.63
Q9HAW8	UGT1A10	0.00009090	−0.95
P01042	KNG1	0.00009660	−0.74
O75356	ENTPD5	0.00012385	−0.54
Q15293	RCN1	0.00012867	0.61
P17931	LGALS3	0.00013151	−0.62
P36952	SERPINB5	0.00013151	1.31

TABLE III—*continued*

UniProt accession number[a]	Gene name	q value	Average fold change (log2)
P09211[b]	GSTP1	0.00014809	0.55
P20774	OGN	0.00015467	−1.52
Q8NBJ4	GOLM1	0.00015467	−0.51
Q16563	SYPL1	0.00015568	1.16
P04229; P13761; Q30134; Q9TQE0; Q9GIY3; Q29974; P04440; P79483	HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DRB1; HLA-DPB1; HLA-DRB3	0.00015568	−0.81
P35555	FBN1	0.00017230	−0.74
Q9H4M9; **Q9NZN4**; **Q9H223**	EHD1; EHD2; EHD4	0.00017230	−0.56
O00754	MAN2B1	0.00017278	−0.65
B4DR31	DPYSL2	0.00018023	−0.55
P06703[b]	S100A6	0.00018025	1.18
P19801	ABP1	0.00022289	−0.50
P20042	EIF2S2	0.00022417	0.68
P51858	HDGF	0.00024151	0.56
P49959[b]	MRE11A	0.00024174	0.64
O00299[b]; **Q9Y696**	CLIC1; CLIC4	0.00024726	0.67
P61604[b]	HSPE1	0.00024977	0.52
O43252; O95340	PAPSS1; PAPSS2	0.00027464	−0.57
Q12765	SCRN1	0.00029349	−0.64
Q9HB40	SCPEP1	0.00030328	−0.87
P48556	PSMD8	0.00032019	0.56
O60547	GMDS	0.00034711	0.58
Q9Y6R7	FCGBP	0.00035167	−0.73
P61626	LYZ	0.00036558	0.74
Q9Y224	C14orf166	0.00038868	0.52
P01765; P01764	(Ig heavy chain V-III region TIL; Ig heavy chain V-III region VH26)	0.00038868	−0.69
Q9NVP1	DDX18	0.00038996	0.67
P80365	HSD11B2	0.00039413	−0.68
P39687; **Q92688**	ANP32A; ANP32B	0.00039654	0.56
Q86WA6	BPHL	0.00046215	0.52
P24298	GPT	0.00047338	−0.56
Q12874	SF3A3	0.00052166	0.50
P04899; **Q14344**; **P63092**	GNAI2; GNA13; GNAS	0.00062192	−0.51
Q15124	PGM5	0.00068593	−0.68
Q9HAW7; **Q9HAW9**; **O60656**	UGT1A7; UGT1A8; UGT1A9	0.00071202	−0.87
P19224	UGT1A6	0.00071202	−0.87
P06748	NPM1	0.00071508	0.81
Q9NUV9	GIMAP4	0.00071508	−0.89
P18283[b]	GPX2	0.00071508	0.54
P13688	CEACAM1	0.00072482	−1.06
P01591	IGJ	0.00084082	−1.10
P19823	ITIH2	0.00085565	−0.71
P01774; P01776; P01779	(Ig heavy chain V-III region POM; Ig heavy chain V-III region WAS; Ig heavy chain V-III region TUR)	0.00085565	−0.79
P27695[b]	APEX1	0.00086971	0.54
Q9C002	NMES1	0.00087016	−0.91
Q96F85	CNRIP1	0.00088956	−1.27
Q9BPX5	ARPC5L	0.00095098	0.67
P62263	RPS14	0.00095182	0.52
Q9BY32	ITPA	0.00095634	0.51
P01625	(Ig kappa chain V-IV region Len)	0.00097041	−0.61
Q15582	TGFBI	0.00097041	0.90
Q07021[b]	C1QBP	0.00101874	0.76
P00738	HP	0.00102030	−0.61
O15143	ARPC1B	0.00103757	−0.50
Q03154	ACY1	0.00108571	0.60
Q9HCB6	SPON1	0.00115609	−0.89
Q96HE7	ERO1L	0.00119027	0.50
P08575	PTPRC	0.00119950	−0.52

TABLE III—*continued*

UniProt accession number[a]	Gene name	q value	Average fold change (log2)
Q9Y266	NUDC	0.00152358	0.56
P63313; P62328	TMSB10; TMSB4X	0.00159602	0.98
Q96SQ9	CYP2S1	0.00162405	0.84
Q71U36	TUBA1A	0.00162405	0.54
P00915	CA1	0.00163258	−1.18
P04844	RPN2	0.00165748	0.55
P09669	COX6C	0.00171230	−0.61
P21980	TGM2	0.00174251	−0.55
P00325	ADH1B	0.00175658	−1.18
O14745	SLC9A3R1	0.00175658	−0.51
Q9H8H3	METTL7A	0.00179938	−0.50
P61009	SPCS3	0.00186488	−0.69
Q15746; O15264; Q16539	MYLK; MAPK13; MAPK14	0.00186488	−0.53
P01876; P01877; **Q92973**	IGHA1; IGHA2; TNPO1	0.00191760	−1.01
P12109	COL6A1	0.00194792	−0.67
Q9BX66	SORBS1	0.00205902	−0.64
E9PGJ9	CC2D1A	0.00213982	−0.53
P49006	MARCKSL1	0.00233532	0.51
Q01524	DEFA6	0.00233532	1.60
P01620	(Ig kappa chain V-III region SIE)	0.00238585	−0.56
P36873	PPP1CC	0.00240171	0.58
Q07507	DPT	0.00240576	−1.28
P37840	SNCA	0.00261720	−0.57
P00326; P07327	ADH1C; ADH1A	0.00271753	−1.17
P22105	TNXB	0.00271753	−0.51
O95299	NDUFA10	0.00272901	−0.84
Q9NRP0	OSTC	0.00275806	0.76
P10082	PYY	0.00282902	−1.59
P21810	BGN	0.00283401	−0.66
Q8IV08	PLD3	0.00295758	−0.74
P01857; P01859; P01860; P01861	IGHG1; IGHG2; IGHG3; IGHG4	0.00343621	−0.57
P62330	ARF6	0.00343645	0.84
Q03135	CAV1	0.00346186	−0.71
P22309; P35503; P22310; P35504	UGT1A1; UGT1A3; UGT1A4; UGT1A5	0.00349148	−0.87
Q9NSU2	TREX1	0.00349148	−0.76
Q9UKY7	CDV3	0.00355684	0.79
Q7Z4V5	HDGFRP2	0.00360855	0.65
P07357	C8A	0.00372327	−0.62
Q99757	TXN2	0.00377179	−0.73
P13686	ACP5	0.00385765	−0.80
Q8WWA0	ITLN1	0.00392858	−1.30
P62861	FAU	0.00395014	0.51
P57737	CORO7	0.00401004	−0.83
P10606	COX5B	0.00412793	−0.71
Q9Y259	CHKB	0.00420774	−0.59
Q9Y2J8	PADI2	0.00435864	−0.50
O94919	ENDOD1	0.00451988	−0.80
B9A064; P0CG05; A0M8Q6	IGLL5; IGLC2; IGLC7	0.00451988	−0.56
P20039	HLA-DRB1	0.00459324	−0.82
P63167[b]; **Q96FJ2**	DYNLL1; DYNLL2	0.00468484	0.63
Q9UNN8	PROCR	0.00480932	−0.78
P07099	EPHX1	0.00486543	−0.54
P32322	PYCR1	0.00495977	0.55
Q9P0J0	NDUFA13	0.00534323	−0.60
E7EUF8; **E9PFN5**	EPB41L3; GSTK1	0.00539222	−0.95
O75531	BANF1	0.00560405	0.73
P26447	S100A4	0.00562754	−0.53
Q9NVJ2	ARL8B	0.00562754	0.50
Q8N752	CSNK1A1	0.00562754	0.52
P40616	ARL1	0.00583778	0.60

TABLE III—*continued*

UniProt accession number[a]	Gene name	q value	Average fold change (log2)
Q96GA7	SDSL	0.00583778	−0.82
P01275	GCG	0.00607808	−1.33
P15289	ARSA	0.00633336	−0.57
O75521	ECI2	0.00635218	−0.60
P62158[b]	CALM1; CALM2; CALM3	0.00657472	0.67
P49821	NDUFV1	0.00669319	−0.66
Q15746–5	MYLK	0.00678109	−0.51
Q96BM9	ARL8A	0.00686655	0.54
Q6UX06	OLFM4	0.00696505	1.14
P10153	RNASE2	0.00724902	−0.50
P19075[b]	TSPAN8	0.00837908	0.59
Q8WU39	PACAP	0.00837978	−0.56
P21953	BCKDHB	0.00837978	0.54
O76041	NEBL	0.00837978	0.71
Q9H4G4	GLIPR2	0.00849532	−1.10
P01766; P01767; P01768	(Ig heavy chain V-III region BRO; Ig heavy chain V-III region BUT; Ig heavy chain V-III region CAM)	0.00895132	−0.61
Q9NR56; Q5VZF2	MBNL1; MBNL2	0.00996176	0.55
P27105	STOM	0.01083127	−0.51
P05387	RPLP2	0.01100903	0.62
Q96AB3	ISOC2	0.01164408	−0.51
O43294	TGFB1I1	0.01198321	−0.57
Q08752	PPID	0.01211602	0.55
Q96DG6	CMBL	0.01289211	−0.51
P61619	SEC61A1	0.01375705	0.59
P56381; **Q5VTU8**	ATP5E; ATP5EP2	0.01440856	−0.52
P14174[b]	MIF	0.01488262	0.51
P12110	COL6A2	0.01526347	−0.53
Q14956	GPNMB	0.01546825	−0.63
P46952	HAAO	0.01570996	−0.53
Q86VN1	VPS36	0.01610077	0.67
Q96S52	PIGS	0.01626862	−0.61
P15559[b]	NQO1	0.01626862	0.56
O60575	SPINK4	0.01810104	0.77
P55735	SEC13	0.01827155	0.59
P02452	COL1A1	0.01933726	−1.32
P00403	MT-CO_2	0.02012815	−0.62

[a] Two or more accession numbers: proteins from the same family or isoforms from the same gene. Boldface numbers indicate "epithelial cell signature" proteins (see text).

[b] Designated candidate cancer biomarkers in the Human Protein Atlas database.

were plotted on a heat map, adenomas and normal mucosa samples formed two distinct clusters (Fig. 4*A*). As shown in Fig. 4*B*, tissue expression levels for the 212 dysregulated proteins showed good correlation ($r = 0.74$, $p < 0.001$, 95% confidence interval = 0.67–0.79) with those of mRNAs for the same genes (measured by our group in another set of colorectal adenomas) (26).

Table IV lists the biological processes that were overrepresented in this set of proteins. At the top of this list was xenobiotic metabolism, a process already linked with adenoma formation on the basis of enrichment studies of transcriptomic datasets conducted by our group (32). Three of the dysregulated proteins involved in this process (CYP2S1, NQO1, and GSTP1) displayed up-regulated expression in adenomas, but most were characterized by tumor-related down-regulation (ADH1B, ADH1C/ADH1A, UGT1A9/UGT1A6, UGT1A1/UGT1A4/UGT1A3/UGT1A5, UGT1A7/UGT1A8, UGDH, MAOA, SULT1A3/SULT1A1, PAPSS1/PAPSS2, UGP2). Network-building analysis revealed that all these proteins were linked by subnetworks controlled by cancer-associated transcription factors, such as SP1 or, less frequently, MYC, HIF1A, or TP53 (supplemental Fig. S3). As noted in Table IV, a very similar picture emerged when GO enrichment was also analyzed in a larger set of 621 dysregulated proteins selected with less stringent criteria (q value cutoff ≤ 0.2; average $\log_2$ fold change $\geq \pm 0.5$).

The expression levels of 111 (52%) of the 212 proteins that were differentially expressed in adenomas were also quantified in cell lines (those shown in bold in Table III and referred to hereinafter as the "epithelial cell signature" proteins). Almost half ($n = 51$, 46%) showed directionally similar tumor-related dysregulation in both analyses. Because cell line stud-

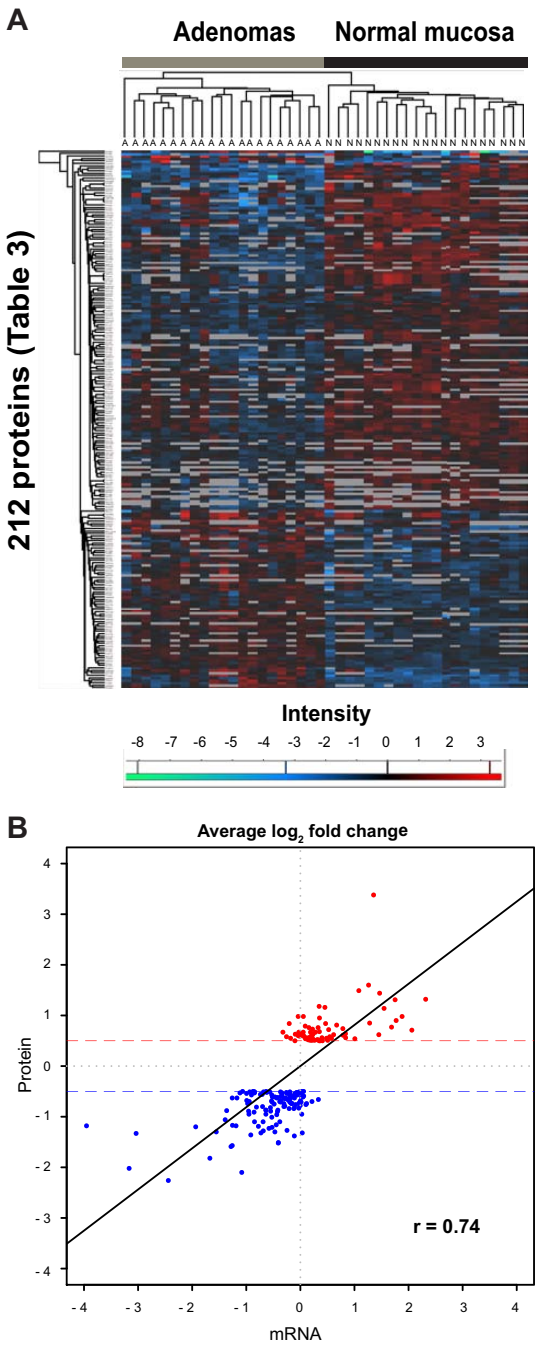


FIG. 4. **Analysis of the 212 proteins displaying significant tumor-related dysregulation.** *A*, hierarchical clustering of iTRAQ abundance ratios (normal *versus* 114, adenoma *versus* 114) for the 212 proteins displaying significant adenoma-related dysregulation grouped tissue samples into two discrete clusters: adenoma (A) and normal (N). *B*, Pearson's correlation test comparing average fold changes (at least ± 0.5 $\log_2$) for the 212 proteins (red, up-regulated; blue, down-regulated) in the tissue series with average $\log_2$ fold changes for the corresponding mRNAs measured in another set of adenoma/normal mucosal samples.

ies were conducted with only one noncancerous line, these findings obviously require further validation. They suggest, however, that these 51 proteins are indeed expressed in the epithelial cells of normal colorectal tissues and that their expression is dysregulated in the epithelial cells of adenomas.

Up-regulation of SORD Expression and Activity in Colorectal Adenomas and Cancer Cell Lines—Sorbitol dehydrogenase, a key enzyme in the polyol pathway, was one of the most significantly up-regulated proteins in our colorectal adenomas (based on q values) (Table III). Because its increased expression could have metabolic consequences with potential effects on tumorigenesis, we performed Western blotting and immunostaining studies to validate this finding. The reliability of the anti-SORD antibody we had chosen was first tested on protein extracts from the six colorectal epithelial cell lines (Fig. 5*A*). The tumor-related $\log_2$ fold changes detected with Western blotting were substantially greater than those documented with iTRAQ (2 to 6 *versus* 0.4 to 1, respectively) (Fig. 5*B*), which was not surprising, as iTRAQ has been reported to underestimate protein abundance (33). However, the relative quantities of SORD found with the two methods were fully consistent. As for the 21 adenomas, the elevated SORD expression documented in these tumors by iTRAQ (Fig. 5*C*) showed good correlation with the increased *SORD* mRNA levels we had previously found in 42 other lesions of this type (26) (Fig. 5*D*). Western blot analysis of four randomly selected adenoma–normal mucosa pairs from the present series revealed obvious up-regulation of SORD expression in all four tumors, although the magnitude of the increase varied (Fig. 5*E*).

SORD activity was then assayed (see "Experimental Procedures") to see how it corresponded with the enzyme expression levels reported above. As shown in Fig. 5*F*, the results of cell line assays were fully consistent with the Western blotting data: SORD activity was seven times higher in HT29 than in HCEC cells, and more limited up-regulation was found in SW480. High correlation between enzyme activity and protein level was also documented for three randomly selected adenoma–normal mucosa pairs (Fig. 5*F*).

MS and Western blotting findings were further validated with immunostaining studies, as shown in Fig. 6. Cytoplasmic SORD staining was evident in the colon cancer cell line HT29 but was weaker or even absent in normal epithelial HCEC cells (Figs. 6*A* and 6*B*). As for colorectal tissues, SORD cytoplasmic expression was limited to the bottom of the normal epithelial crypts (Figs. 6*C*, 6*D*, and 6*E*), but its expression was markedly increased in adenomatous and cancerous glands (Figs. 6*F*–6*I*). These findings suggest that SORD is likely to be

TABLE IV
Gene Ontology (GO) biological processes enriched in the set of 212 proteins whose expression displayed adenoma-related dysregulation (see Table III)

GO I.D.	GO term	Annotated[a]	Significant[b]	Up in adenomas	Down in adenomas	Expected[c]	*elim p* value[d]
GO:0006805	Xenobiotic metabolic process[e]	64	21	3	18	2.8	1.70E-09
GO:0006958	Complement activation, classical pathway[e]	74	17	0	17	3.24	1.10E-08
GO:0051552	Flavone metabolic process[e]	5	5	0	5	0.22	1.50E-07
GO:0052696	Flavonoid glucuronidation[e]	5	5	0	5	0.22	1.50E-07
GO:0052697	Xenobiotic glucuronidation[e]	5	5	0	5	0.22	1.50E-07
GO:0042573	Retinoic acid metabolic process[e]	9	6	0	6	0.39	5.00E-07
GO:0045087	Innate immune response[f]	249	33	4	29	10.91	7.80E-07
GO:0031295	T-cell costimulation[f]	34	10	0	10	1.49	1.10E-06
GO:0030199	Collagen fibril organization[e]	8	5	0	5	0.35	7.80E-06
GO:0001501	Skeletal system development	54	11	1	10	2.37	1.60E-05
GO:0070208	Protein heterotrimerization[e]	5	4	0	4	0.22	1.70E-05
GO:0050852	T-cell receptor signaling pathway[f]	48	10	0	10	2.1	3.20E-05

[a] Proteins in TopGO Background list.
[b] 212 dysregulated proteins of Table III.
[c] Number of significant proteins expected to map to the GO term if the significant proteins are randomly distributed over all GO terms.
[d] *p* value from the Eliminating Genes method (25). Only processes with an *elim p* value $<$ 1.0E-04 are shown.
[e] Processes that were also among the top 12 processes displaying enrichment in a larger set of 621 dysregulated proteins selected with less stringent criteria (q value $\leq$ 0.2; average $\log_2$ fold change $\geq \pm 0.5$; see "Results" section for details).
[f] Processes that shared a common GO ancestor (immune system process) with the process displaying the most significant enrichment in the larger set.

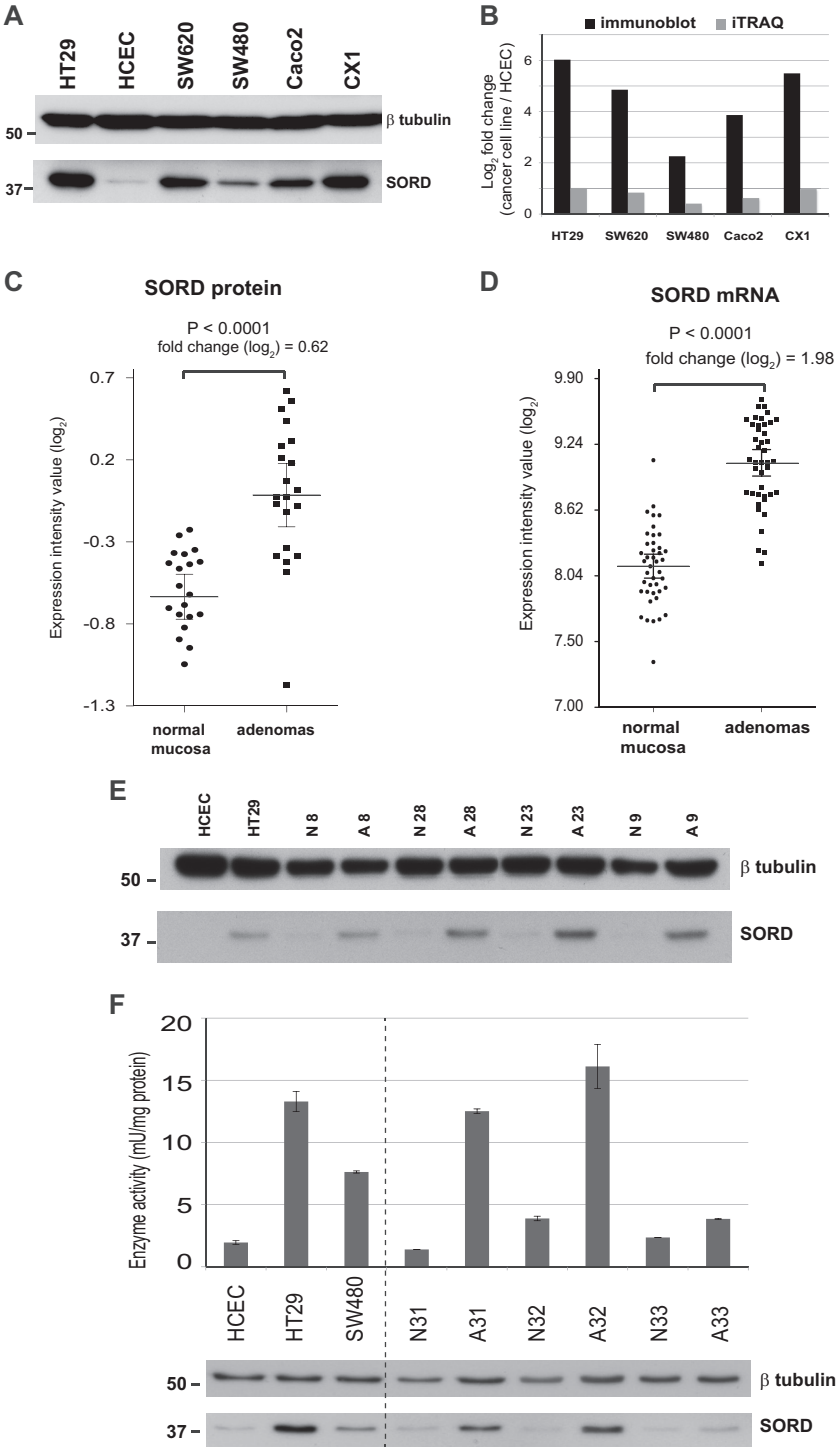
expressed in proliferating cells, although it was largely absent in HCECs, which undergo regular proliferation *in vitro*. Furthermore, nuclear localization of SORD was noted in some adenomatous crypts (supplemental Figs. S4*A* and S4*C*), and the cells in question were almost always negative for the well-known proliferation marker Ki-67 (supplemental Figs. S4*B* and S4*D*). This mutually exclusive staining pattern was also observed in normal crypts of the ileum, where SORD, interestingly, appeared to be expressed in the nuclei of putative stem cells (supplemental Figs. S4*E* and S4*F*).

Polyol Pathway Enzyme Expression and Metabolite Levels in Cell Lines and Tissues—We then examined the state of the polyol pathway (supplemental Fig. S5*A*) in colorectal cell lines and tissues. As shown in supplemental Fig. S5*B*, immunoblot studies revealed decreased AKR1B1 expression in HT29 (*versus* HCEC cells) and adenomas (*versus* corresponding normal mucosal samples), whereas SORD expression and that of KHK were up-regulated in tumor cells and tissues. As for the metabolites (supplemental Fig. S5*C*), D-glucose levels were significantly decreased in adenomas. Less dramatic changes were observed in the levels of D-sorbitol and D-fructose, which both showed a tendency to decrease in tumor tissues.

DISCUSSION

Although a number of proteomic studies have comparatively analyzed different types of colorectal tissues, precancerous lesions have been considered in only three (21, 34, 35), and in two of these (21, 34), the number of adenomas analyzed was very small ($\leq$4). The study by Lam *et al.* (35) is the only one that compared protein expression in a relatively large number (n = 20) of paired adenoma and normal mucosa samples. They used two-dimensional gel electrophoresis to resolve over 1000 proteins in the two tissue groups, and those displaying differential expression were then analyzed with MALDI-TOF/TOF-MS. MS/MS validation pinpointed four proteins (ANXA3, S100A11, EIF-5A1, and S100P) whose expression in adenomas was significantly increased. Using MS with iTRAQ 8-plex peptide labeling and OFFGEL fractionation allowed us to quantitatively compare protein expression in 30 colorectal adenomas and paired samples of normal mucosa and investigate low-abundance proteins that cannot be evaluated with proteomics based on two-dimensional gel electrophoresis. All in all, 4325 nonredundant protein families were quantified in our colorectal tissues (25% of which were identified in all 60 samples) (Table II), and the 212 proteins we flagged as significantly dysregulated in adenomas included three of the four proteins identified by Lam *et al.* (up-regulation of the fourth, EIF-5A1, failed to meet our stringent criterion for significance) (Table III).

The cell types in which these proteomic changes occur is of obvious interest, as colorectal cancer arises from the epithelial component of the colorectal mucosa. Although our findings are preliminary and will naturally require validation in future studies, 51 of the 212 proteins listed in Table III were "epithelial cell signature" proteins and showed directionally similar expression changes in colon cancer cell lines *versus* HCEC. It therefore seems likely that their dysregulated expression in adenomas is a feature of neoplastic transformation of colorectal epithelial cells. However, epithelial–stromal cell interactions can also play important roles in tumorigenesis (20). Our approach also allowed us to identify 101 proteins displaying adenoma-related dysregulation that were probably


A
HT29
HCEC
SW620
SW480
Caco2
CX1
50
β tubulin
37
SORD
B
immunoblot
iTRAQ
Log2 fold change (cancer cell line / HCEC)
HT29
SW620
SW480
Caco2
CX1
C
SORD protein
P < 0.0001
fold change (log2) = 0.62
Expression intensity value (log2)
normal mucosa
adenomas
D
SORD mRNA
P < 0.0001
fold change (log2) = 1.98
Expression intensity value (log2)
normal mucosa
adenomas
E
HCEC
HT29
N 8
A 8
N 28
A 28
N 23
A 23
N 9
A 9
β tubulin
SORD
F
Enzyme activity (mU/mg protein)
HCEC
HT29
SW480
N31
A31
N32
A32
N33
A33
β tubulin
SORD

of stromal-cell origin, as they were not expressed in any of the six epithelial cell lines we examined (Table III). These proteins were mainly involved in immune-related processes (immune response, complement activation, T-cell co-stimulation), which are usually not represented in colon epithelial cell lines. Their expression changes are likely to have important effects on the microenvironment of an epithelial-cell tumor.

Our search for potential biomarkers of early-stage colorectal tumorigenesis focused exclusively on the 76 proteins whose expression was significantly up-regulated in adenomas. According to the Human Protein Atlas database (36), 69 (91%) of these have cancer-related features, and 16 of the 69 are already classified as candidate cancer biomarkers (Table III). The Human Protein Atlas database contains information on protein expression in normal and cancer tissues, but not in those regarded as precancerous. The overlap between our findings and those of the Human Protein Atlas suggest that most protein expression changes identified thus far in colorectal adenocarcinomas are probably already detectable in the benign precursors of these lesions. Supplemental Fig. S6 shows the expression profiles of the 10 proteins that were most markedly up-regulated in adenomas. This group included two of the four proteins identified by Lam *et al.* (35) as significantly overexpressed in adenomas. Annexin A3 (ANXA3), for example, is at the top of our list (based on q values) (Table III). An angiogenic factor that induces VEGF production via the HIF-1 pathway (37), ANXA3 belongs to a family of calcium-dependent, phospholipid-binding proteins involved in diverse biological processes, including signal transduction, inflammatory responses, membrane organization, and the regulation of cellular growth (38, 39). Dysregulated ANXA expression is also a common feature of colorectal cancer (39), and most other cancers as well (40). S100A11 expression was also increased in these tumors, which is consistent with earlier reports (41). The cytosolic S100 proteins interact directly with peptides on the N-terminal domain of annexins (38, 42), and like the annexins, they also have diverse intracellular and extracellular functional roles (43).

Among the other top 10 proteins displaying adenoma-related up-regulation were LDHA and LDHB. Their expression levels were not measured separately, but LDHA is presumably responsible for the increased expression observed in our adenomas. *LDHB* expression is in most cases epigenetically silenced in colon cancer cells (44, 45), whereas *LDHA* is overexpressed, and its activity is maintained via the oncogenic tyrosine kinase FGFR1 (46). LDHA is a key player in the reversible conversion of pyruvate to lactate during aerobic glycolysis, a typical feature of cancer cell metabolism first described by Warburg (47). The sodium- and potassium-coupled chloride cotransporter SLC12A2 is expressed on the basolateral membrane of the normal colon epithelium, where its recruitment and activation are regulated by calcium and cAMP. Loss of SLC12A2 leads to impaired chloride secretion in the intestine (48, 49), but to our knowledge, there are no published data linking this protein to colon cancer. The fifth markedly overexpressed protein was SET, one of the five proteins that make up the inhibitor of acetyltranferases (INHAT) complex. Two other INHAT components, APEX1 and ANP32A/ANP32B, were also up-regulated in adenomas (albeit to a lesser extent than SET) (Table III). These changes are noteworthy because INHAT binds directly to histones, preventing their acetylation by histone acetyl transferases (50–52), and loss of histone acetylation is a crucial step in gene silencing (53, 54). Thus far, INHAT's role in cancer has not been widely investigated, but overexpression of the complex components has been observed in serous epithelial ovarian cancer (55). The up-regulated expression of PPA1/PPA2 in our adenomas might play various roles in colorectal tumorigenesis, as these proteins are key players in the synthesis of fatty acids, nucleotides, amino acids, and other essential molecules (56). The phosphoprotein nucleolin, an essential protein for proliferating cells (57), appears to regulate several steps in the biogenesis of ribosomes, including transcription, ribosome assembly, and the processing of precursor ribosomal RNA (58–60), all of which might be instrumental in adenoma growth. As for OCIAD2, strong immunoreactivity for this protein has been reported in early-stage adenocarcinomas of the lung and in ovarian cancers (61–63), but there are no published data linking it to colorectal tumorigenesis. In contrast, the secreted protein REG4, which promotes mitosis and enhances the motility and invasiveness of colon cancer cells, is strongly expressed in these cells and in the serum of patients with colorectal cancer (64–66).

The final protein characterized by marked adenoma-related up-regulation was SORD, a key enzyme in the polyol metabolic pathway. It was selected for validation studies because although aberrant polyol pathway activity has been implicated in diabetic complications (67–70) and myocardial ischemia

FIG. 5. **Significantly up-regulated SORD expression and activity in colorectal cell lines and adenomas.** *A*, tumor-related up-regulation of sorbitol dehydrogenase (SORD) in colon cancer cell lines was confirmed with Western blotting. The SORD dysregulation trend was identical to that observed with iTRAQ-based MS/MS, although when immunoblot results were quantified (*B*), the $\log_2$ fold changes were more than five times greater than those documented in the iTRAQ study. *C*, SORD protein expression (iTRAQ analysis) in 21 normal mucosa–adenoma tissue pairs. *D*, *SORD* mRNA expression in 42 other normal mucosa–adenoma pairs from a previous study by our group (26). Error bars indicate the means and 95% confidence intervals. *E*, Western blots showing tumor-related up-regulation of SORD expression in four randomly selected adenoma (A)/normal mucosa (N) tissue pairs of the 21 shown in panel *C* (see Table I for sample descriptions). *F*, SORD activity also displayed tumor-related up-regulation in cell lines (HT29 and SW480 *versus* HCEC cells) and tissues (adenomas *versus* normal mucosa). Columns show mean enzyme activity measured in at least two replicates; error bars indicate standard deviations from means. The Western blot beneath the graph shows SORD levels measured in the extracts used for the enzyme activity assays.

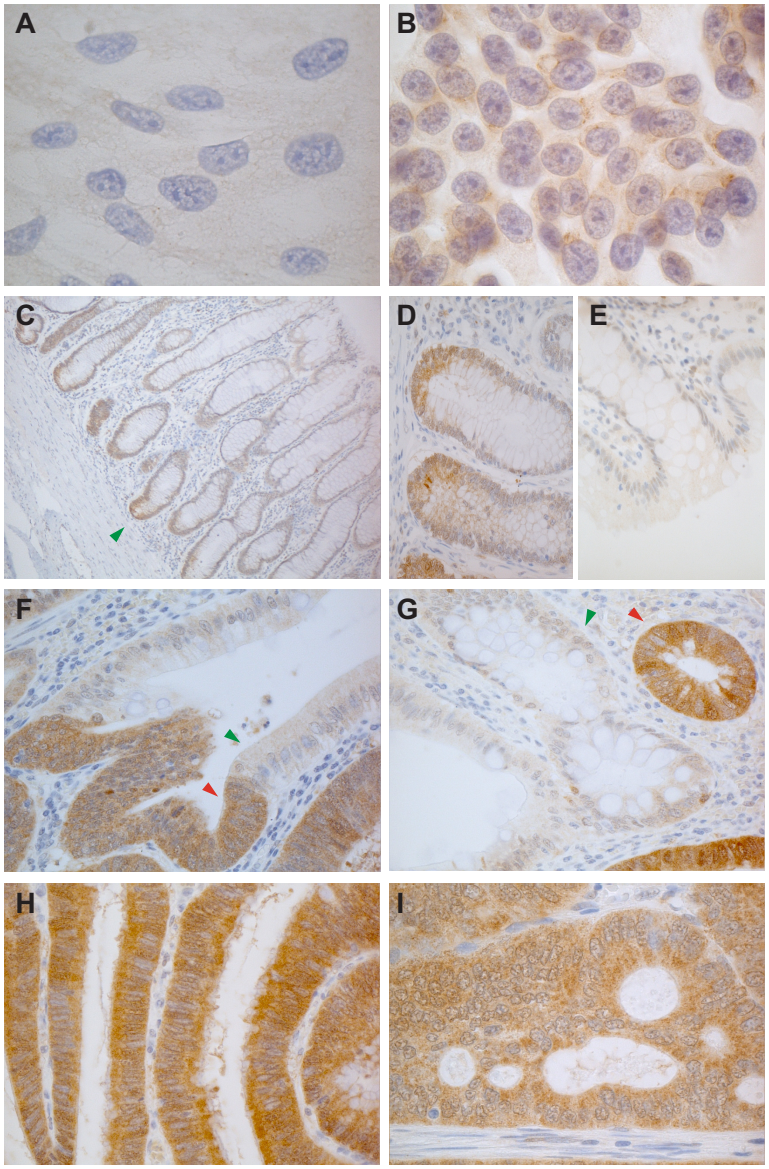


FIG. 6. **Anti-SORD immunostaining of colorectal cell lines and tissues.** Consistent with proteomic data, SORD expression was (*A*) negligible or absent in HCECs but (*B*) clearly expressed in the cytoplasm of HT29 cells. *C*, in normal colorectal mucosa, SORD expression was limited to the lower portion of the epithelial crypts, where stem cells and highly proliferating cells are located. Higher magnification views show staining at (*D*) the base *versus* (*E*) the mouth of colonic crypts. *F*, *G*, its expression was markedly increased in adenomatous glands (red arrowheads) relative to normal crypts (green arrowheads). Panels *H* and *I* show abundant expression of SORD in a large adenoma and in a cancer, respectively.

(71), the role of SORD in tumorigenesis was completely unknown. During the execution of this study, however, up-regulated SORD expression was reported in prostate cancer (72) and in colorectal adenomas (21), and these findings strengthened our resolve to characterize this phenomenon in colorectal tumorigenesis.

Up-regulated SORD expression and activity in adenomas (Fig. 5) would enhance the production of fructose (see sche-

matic of supplemental Fig. S5*A*), thereby increasing the generation of triose sugars and diacylglycerol (intermediates in the glycolytic and lipid signaling pathways, respectively). Fructose is also several times more effective than glucose in promoting intracellular non-enzymatic glycation (73–75), and advanced glycation end products may contribute to the vascular complications of diabetes and other pathologic conditions (67, 76–78). Whether these fructose-driven metabolic events play a role in the development of adenomas is unclear, but the polyol pathway was very active in the adenomatous cells we examined. This activity was also reflected in the concomitant increase of the expression of KHK (supplemental Fig. S5*B*), the enzyme that catalyzes the transformation of fructose to fructose-1-P, downstream from the polyol pathway.

The effects of these enzymatic changes on sorbitol and fructose concentrations in adenomas need to be investigated in larger tissue series, but our preliminary data suggest that the levels of both are slightly decreased in these lesions (supplemental Fig. S5*C*). In contrast, our adenomas exhibited dramatically reduced concentrations of glucose, the initial substrate in the polyol pathway (supplemental Fig. S5*C*). Adenoma-related dysregulation was also noted in the expression of AKR1B1, the enzyme that converts glucose to sorbitol (supplemental Fig. S5*B*). Exploitation of the polyol pathway to divert carbon from glucose to other energy intermediates might provide adenomatous cells with a selective advantage over normal cells. This pathway might prove to be another means of tumor-related glucose consumption in addition to the well-known glycolytic and pentose phosphate pathways (supplemental Fig. S5*A*). Advanced cancer cells consume glucose at a much higher rate than normal cells, and much of their energy is generated by aerobic glycolysis rather than by oxidative phosphorylation of glucose in the mitochondria (*i.e.* the Warburg effect) (79). The predominantly glycolytic phenotype of cancer cells results in low glucose levels and high concentrations of lactate (47, 80, 81). The relative concentrations of lactate in the three adenomas we tested were significantly greater than those found in matched samples of normal mucosa (supplemental Fig. S5*C*), indicating that the Warburg effect is already evident in precancerous colorectal lesions. Studies involving metabolic flux analysis to monitor the fate of isotopic tracers in *in vitro* and *in vivo* systems would provide further insight into the biological roles of the polyol pathway in tumorigenesis. Further information on selected PSMs, peptides, and corresponding assembled proteins can be found in supplemental Tables S1 (tissues) and S3 (cell lines). Supplemental Figs. S7 through S17 show spectra for the proteins identified with a single peptide (listed in supplemental Tables S1 and S3).

Acknowledgments—We thank Ritva Haider, Mirjam Schneider, Asa Wahlander, Bernd Roschitzki, Claudia Fortes, Tshering Altherr, Michal J. Okoniewski, and Katrin Hecht for technical assistance; Paola Picotti, Bernd Wollscheid, Joseph Jiricny, Javier Pena Diaz, Mirco Menigatti, and David Fischer for productive discussion; and Marian Everett Kent for editing the manuscript. The MS data have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD000445 and DOI 10.6019/PXD000445.

* The research was supported by a grant from the Swiss National Science Foundation (Grant No. 31003A_141225).

S This article contains supplemental material.

** To whom correspondence should be addressed: Giancarlo Marra, M.D., Ph.D., Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; Tel.: 0041-44-635-3472; Fax: 0041-44-635-3484; E-mail: marra@imcr.uzh.ch.

REFERENCES

1. Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011) Global cancer statistics. *CA Cancer J. Clin.* **61,** 69–90
2. Siegel, R., Naishadham, D., and Jemal, A. (2013) Cancer statistics, 2013. *CA Cancer J. Clin.* **63,** 11–30
3. Shinya, H., and Wolff, W. I. (1979) Morphology, anatomic distribution and cancer potential of colonic polyps. *Ann. Surg.* **190,** 679–683
4. Cattaneo, E., Baudis, M., Buffoli, F., Bianco, M. A., Zorzi, F., and Marra, G. (2011) Pathways and crossroads to colorectal cancer In *Pre-Invasive Disease: Pathogenesis and Clinical Management* (Fitzgerald, R. C., ed), pp. 369–394, Springer, New York
5. U.S. Multi-Society Task Force (2002) Screening for colorectal cancer: recommendation and rationale. *Ann. Intern. Med.* **137,** 129–131
6. Kahi, C. J., Rex, D. K., and Imperiale, T. F. (2008) Screening, surveillance, and primary prevention for colorectal cancer: a review of the recent literature. *Gastroenterology* **135,** 380–399
7. Levin, B., Lieberman, D. A., McFarland, B., Andrews, K. S., Brooks, D., Bond, J., Dash, C., Giardiello, F. M., Glick, S., Johnson, D., Johnson, C. D., Levin, T. R., Pickhardt, P. J., Rex, D. K., Smith, R. A., Thorson, A., Winawer, S. J., American Cancer Society Colorectal Cancer Advisory, U.S. Multi-Society Task Force, and American College of Radiology Colon Cancer (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology* **134,** 1570–1595
8. Sillars-Hardebol, A. H., Carvalho, B., van Engeland, M., Fijneman, R. J., and Meijer, G. A. (2012) The adenoma hunt in colorectal cancer screening: defining the target. *J. Pathol.* **226,** 1–6
9. de Wit, M., Fijneman, R. J., Verheul, H. M., Meijer, G. A., and Jimenez, C. R. (2013) Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. *Clin. Biochem.* **46,** 466–479
10. Patterson, S. D., and Aebersold, R. H. (2003) Proteomics: the first decade and beyond. *Nat. Genet.* **33 Suppl,** 311–323
11. Horth, P., Miller, C. A., Preckel, T., and Wenz, C. (2006) Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol. Cell. Proteomics* **5,** 1968–1974
12. Wu, W. W., Wang, G., Baek, S. J., and Shen, R. F. (2006) Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel or LC-MALDI TOF/TOF. *J. Proteome Res.* **5,** 651–658
13. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
14. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics. *Chem. Rev.* **101,** 269–295
15. Bensimon, A., Heck, A. J., and Aebersold, R. (2012) Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **81,** 379–405
16. Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404,** 939–965
17. Wiese, S., Reidegeld, K. A., Meyer, H. E., and Warscheid, B. (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7,** 340–350
18. Gan, C. S., Chong, P. K., Pham, T. K., and Wright, P. C. (2007) Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. Proteome Res.* **6,** 821–827

19. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3,** 1154–1169
20. Jimenez, C. R., Knol, J. C., Meijer, G. A., and Fijneman, R. J. (2010) Proteomics of colorectal cancer: overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *J. Proteomics* **73,** 1873–1895
21. Besson, D., Pavageau, A. H., Valo, I., Bourreau, A., Belanger, A., Eymerit-Morin, C., Mouliere, A., Chassevent, A., Boisdron-Celle, M., Morel, A., Solassol, J., Campone, M., Gamelin, E., Barre, B., Coqueret, O., and Guette, C. (2011) A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. *Mol. Cell. Proteomics* **10,** M111.009712
22. Jankova, L., Chan, C., Fung, C. L., Song, X., Kwun, S. Y., Cowley, M. J., Kaplan, W., Dent, O. F., Bokey, E. L., Chapuis, P. H., Baker, M. S., Robertson, G. R., Clarke, S. J., and Molloy, M. P. (2011) Proteomic comparison of colorectal tumours and non-neoplastic mucosa from paired patient samples using iTRAQ mass spectrometry. *Mol. Biosyst.* **7,** 2997–3005
23. Roig, A. I., Eskiocak, U., Hight, S. K., Kim, S. B., Delgado, O., Souza, R. F., Spechler, S. J., Wright, W. E., and Shay, J. W. (2010) Immortalized epithelial cells derived from human colon biopsies express stem cell markers and differentiate in vitro. *Gastroenterology* **138,** 1012–1021
24. Kall, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7,** 40–44
25. Alexa, A., Rahnenfuhrer, J., and Lengauer, T. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22,** 1600–1607
26. Cattaneo, E., Laczko, E., Buffoli, F., Zorzi, F., Bianco, M. A., Menigatti, M., Bartosova, Z., Haider, R., Helmchen, B., Sabates-Bellver, J., Tiwari, A., Jiricny, J., and Marra, G. (2011) Preinvasive colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid vs. nonpolypoid). *EMBO Mol. Med.* **3,** 334–347
27. Truninger, K., Menigatti, M., Luz, J., Russell, A., Haider, R., Gebbers, J. O., Bannwart, F., Yurtsever, H., Neuweiler, J., Riehle, H. M., Cattaruzza, M. S., Heinimann, K., Schär, P., Jiricny, J., and Marra, G. (2005) Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. *Gastroenterology* **128,** 1160–1171
28. Gerlach, U., and Hilby, W. (1974) *Methods of Enzymatic Analysis*, 2nd Ed., Academic Press Inc., New York
29. Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A. R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1,** 387–396
30. Afkarian, M., Bhasin, M., Dillon, S. T., Guerrero, M. C., Nelson, R. G., Knowler, W. C., Thadhani, R., and Libermann, T. A. (2010) Optimizing a proteomics platform for urine biomarker discovery. *Mol. Cell. Proteomics* **9,** 2195–2204
31. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4,** 1265–1272
32. Maglietta, R., Liuzzi, V. C., Cattaneo, E., Laczko, E., Piepoli, A., Panza, A., Carella, M., Palumbo, O., Staiano, T., Buffoli, F., Andriulli, A., Marra, G., and Ancona, N. (2012) Molecular pathways undergoing dramatic transcriptomic changes during tumor development in the human colon. *BMC Cancer* **12,** 608
33. Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. (2009) iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly." *J. Proteome Res.* **8,** 5347–5355
34. Albrethsen, J., Knol, J. C., Piersma, S. R., Pham, T. V., de Wit, M., Mongera, S., Carvalho, B., Verheul, H. M., Fijneman, R. J., Meijer, G. A., and Jimenez, C. R. (2010) Subnuclear proteomics in colorectal cancer: identification of proteins enriched in the nuclear matrix fraction and regulation in adenoma to carcinoma progression. *Mol. Cell. Proteomics* **9,** 988–1005
35. Lam, F., Jankova, L., Dent, O. F., Molloy, M. P., Kwun, S. Y., Clarke, C., Chapuis, P., Robertson, G., Beale, P., Clarke, S., Bokey, E. L., and Chan, C. (2010) Identification of distinctive protein expression patterns in colorectal adenoma. *Proteomics Clin. Appl.* **4,** 60–70
36. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Bjorling, L., and Ponten, F. (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28,** 1248–1250
37. Park, J. E., Lee, D. H., Lee, J. A., Park, S. G., Kim, N. S., Park, B. C., and Cho, S. (2005) Annexin A3 is a potential angiogenic mediator. *Biochem. Biophys. Res. Commun.* **337,** 1283–1287
38. Gerke, V., Creutz, C. E., and Moss, S. E. (2005) Annexins: linking Ca2+ signalling to membrane dynamics. *Nat. Rev. Mol. Cell Biol.* **6,** 449–461
39. Duncan, R., Carpenter, B., Main, L. C., Telfer, C., and Murray, G. I. (2008) Characterisation and protein expression profiling of annexins in colorectal cancer. *Br. J. Cancer* **98,** 426–433
40. Mussunoor, S., and Murray, G. I. (2008) The role of annexins in tumour development and progression. *J. Pathol.* **216,** 131–140
41. Melle, C., Ernst, G., Schimmel, B., Bleul, A., Mothes, H., Kaufmann, R., Settmacher, U., and Von Eggeling, F. (2006) Different expression of calgizzarin (S100A11) in normal colonic epithelium, adenoma and colorectal carcinoma. *Int. J. Oncol.* **28,** 195–200
42. Lewit-Bentley, A., Rety, S., Sopkova-de Oliveira Santos, J., and Gerke, V. (2000) S100-annexin complexes: some insights from structural studies. *Cell Biol. Int.* **24,** 799–802
43. Donato, R. (2001) S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int. J. Biochem. Cell Biol.* **33,** 637–668
44. Thangaraju, M., Carswell, K. N., Prasad, P. D., and Ganapathy, V. (2009) Colon cancer cells maintain low levels of pyruvate to avoid cell death caused by inhibition of HDAC1/HDAC3. *Biochem. J.* **417,** 379–389
45. Thorn, C. C., Freeman, T. C., Scott, N., Guillou, P. J., and Jayne, D. G. (2009) Laser microdissection expression profiling of marginal edges of colorectal tumours reveals evidence of increased lactate metabolism in the aggressive phenotype. *Gut* **58,** 404–412
46. Fan, J., Hitosugi, T., Chung, T. W., Xie, J., Ge, Q., Gu, T. L., Polakiewicz, R. D., Chen, G. Z., Boggon, T. J., Lonial, S., Khuri, F. R., Kang, S., and Chen, J. (2011) Tyrosine phosphorylation of lactate dehydrogenase A is important for NADH/NAD(+) redox homeostasis in cancer cells. *Mol. Cell. Biol.* **31,** 4938–4950
47. Koppenol, W. H., Bounds, P. L., and Dang, C. V. (2011) Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer* **11,** 325–337
48. Reynolds, A., Parris, A., Evans, L. A., Lindqvist, S., Sharp, P., Lewis, M., Tighe, R., and Williams, M. R. (2007) Dynamic and differential regulation of NKCC1 by calcium and cAMP in the native human colonic epithelium. *J. Physiol.* **582,** 507–524
49. Flagella, M., Clarke, L. L., Miller, M. L., Erway, L. C., Giannella, R. A., Andringa, A., Gawenis, L. R., Kramer, J., Duffy, J. J., Doetschman, T., Lorenz, J. N., Yamoah, E. N., Cardell, E. L., and Shull, G. E. (1999) Mice lacking the basolateral Na-K-2Cl cotransporter have impaired epithelial chloride secretion and are profoundly deaf. *J. Biol. Chem.* **274,** 26946–26955
50. Schneider, R., Bannister, A. J., Weise, C., and Kouzarides, T. (2004) Direct binding of INHAT to H3 tails disrupted by modifications. *J. Biol. Chem.* **279,** 23859–23862
51. Kutney, S. N., Hong, R., Macfarlan, T., and Chakravarti, D. (2004) A signaling role of histone-binding proteins and INHAT subunits pp32 and Set/TAF-Ibeta in integrating chromatin hypoacetylation and transcriptional repression. *J. Biol. Chem.* **279,** 30850–30855
52. Seo, S. B., McNamara, P., Heo, S., Turner, A., Lane, W. S., and Chakravarti, D. (2001) Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell* **104,** 119–130
53. Kondo, Y., and Issa, J. P. (2004) Epigenetic changes in colorectal cancer. *Cancer Metastasis Rev.* **23,** 29–39
54. Kondo, Y., Shen, L., and Issa, J. P. (2003) Critical role of histone methylation in tumor suppressor gene silencing in colorectal cancer. *Mol. Cell. Biol.* **23,** 206–215
55. Ouellet, V., Le Page, C., Guyot, M. C., Lussier, C., Tonin, P. N., Provencher, D. M., and Mes-Masson, A. M. (2006) SET complex in serous epithelial ovarian cancer. *Int. J. Cancer* **119,** 2119–2126

56. Takahashi, K., Inuzuka, M., and Ingi, T. (2004) Cellular signaling mediated by calphoglin-induced activation of IPP and PGM. *Biochem. Biophys. Res. Commun.* **325,** 203–214
57. Lapeyre, B., Bourbon, H., and Amalric, F. (1987) Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *Proc. Natl. Acad. Sci. U.S.A.* **84,** 1472–1476
58. Ginisty, H., Amalric, F., and Bouvet, P. (1998) Nucleolin functions in the first step of ribosomal RNA processing. *EMBO J.* **17,** 1476–1486
59. Ginisty, H., Sicard, H., Roger, B., and Bouvet, P. (1999) Structure and functions of nucleolin. *J. Cell Sci.* **112 (Pt 6),** 761–772
60. Tajrishi, M. M., Tuteja, R., and Tuteja, N. (2011) Nucleolin: the most abundant multifunctional phosphoprotein of nucleolus. *Commun. Integr. Biol.* **4,** 267–275
61. Chien, J., Fan, J. B., Bell, D. A., April, C., Klotzle, B., Ota, T., Lingle, W. L., Gonzalez Bosquet, J., Shridhar, V., and Hartmann, L. C. (2009) Analysis of gene expression in stage I serous tumors identifies critical pathways altered in ovarian cancer. *Gynecol. Oncol.* **114,** 3–11
62. Ishiyama, T., Kano, J., Anami, Y., Onuki, T., Iijima, T., Morisita, Y., Yokota, J., and Noguchi, M. (2007) OCIA domain containing 2 is highly expressed in adenocarcinoma mixed subtype with bronchioloalveolar carcinoma component and is associated with better prognosis. *Cancer Sci.* **98,** 50–57
63. Nagata, C., Kobayashi, H., Sakata, A., Satomi, K., Minami, Y., Morishita, Y., Ohara, R., Yoshikawa, H., Arai, Y., Nishida, M., and Noguchi, M. (2012) Increased expression of OCIA domain containing 2 during stepwise progression of ovarian mucinous tumor. *Pathol. Int.* **62,** 471–476
64. Oue, N., Kuniyasu, H., Noguchi, T., Sentani, K., Ito, M., Tanaka, S., Setoyama, T., Sakakura, C., Natsugoe, S., and Yasui, W. (2007) Serum concentration of Reg IV in patients with colorectal cancer: overexpression and high serum levels of Reg IV are associated with liver metastasis. *Oncology* **72,** 371–380
65. Rafa, L., Dessein, A. F., Devisme, L., Buob, D., Truant, S., Porchet, N., Huet, G., Buisine, M. P., and Lesuffleur, T. (2010) REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. *Int. J. Oncol.* **36,** 689–698
66. Violette, S., Festor, E., Pandrea-Vasile, I., Mitchell, V., Adida, C., Dussaulx, E., Lacorte, J. M., Chambaz, J., Lacasa, M., and Lesuffleur, T. (2003) Reg IV, a new member of the regenerating gene family, is overexpressed in colorectal carcinomas. *Int. J. Cancer* **103,** 185–193
67. Obrosova, I. G. (2005) Increased sorbitol pathway activity generates oxidative stress in tissue sites for diabetic complications. *Antioxid. Redox Signal.* **7,** 1543–1552
68. Ola, M. S., Berkich, D. A., Xu, Y., King, M. T., Gardner, T. W., Simpson, I., and LaNoue, K. F. (2006) Analysis of glucose metabolism in diabetic rat retinas. *Am. J. Physiol. Endocrinol. Metab.* **290,** E1057-E1067
69. Van den Enden, M. K., Nyengaard, J. R., Ostrow, E., Burgan, J. H., and Williamson, J. R. (1995) Elevated glucose levels increase retinal glycolysis and sorbitol pathway metabolism. Implications for diabetic retinopathy. *Invest. Ophthalmol. Vis. Sci.* **36,** 1675–1685
70. Gabbay, K. H., Merola, L. O., and Field, R. A. (1966) Sorbitol pathway: presence in nerve and cord with substrate accumulation in diabetes. *Science* **151,** 209–210
71. Hwang, Y. C., Bakr, S., Ellery, C. A., Oates, P. J., and Ramasamy, R. (2003) Sorbitol dehydrogenase: a novel target for adjunctive protection of ischemic myocardium. *FASEB J.* **17,** 2331–2333
72. Szabo, Z., Hamalainen, J., Loikkanen, I., Moilanen, A. M., Hirvikoski, P., Vaisanen, T., Paavonen, T. K., and Vaarala, M. H. (2010) Sorbitol dehydrogenase expression is regulated by androgens in the human prostate. *Oncol. Rep.* **23,** 1233–1239
73. Obrosova, I. G. (2005) Increased sorbitol pathway activity generates oxidative stress in tissue sites for diabetic complications. *Antioxid. Redox Signal.* **7,** 1543–1552
74. Dills, W. L., Jr. (1993) Protein fructosylation: fructose and the Maillard reaction. *Am. J. Clin. Nutr.* **58,** 779S-787S
75. Suarez, G., Rajaram, R., Oronsky, A. L., and Gawinowicz, M. A. (1989) Nonenzymatic glycation of bovine serum albumin by fructose (fructation). Comparison with the Maillard reaction initiated by glucose. *J. Biol. Chem.* **264,** 3674–3679
76. Schalkwijk, C. G., Stehouwer, C. D., and van Hinsbergh, V. W. (2004) Fructose-mediated non-enzymatic glycation: sweet coupling or bad modification. *Diabetes Metab. Res. Rev.* **20,** 369–382
77. Bose, T., and Chakraborti, A. S. (2008) Fructose-induced structural and functional modifications of hemoglobin: implication for oxidative stress in diabetes mellitus. *Biochim. Biophys. Acta* **1780,** 800–808
78. Schmidt, A. M., Hori, O., Brett, J., Yan, S. D., Wautier, J. L., and Stern, D. (1994) Cellular receptors for advanced glycation end products. Implications for induction of oxidant stress and cellular dysfunction in the pathogenesis of vascular lesions. *Arterioscler. Thromb.* **14,** 1521–1528
79. Warburg, O. (1956) On the origin of cancer cells. *Science* **123,** 309–314
80. Chan, E. C., Koh, P. K., Mal, M., Cheah, P. Y., Eu, K. W., Backshall, A., Cavill, R., Nicholson, J. K., and Keun, H. C. (2009) Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J. Proteome Res.* **8,** 352–361
81. Hirayama, A., Kami, K., Sugimoto, M., Sugawara, M., Toki, N., Onozuka, H., Kinoshita, T., Saito, N., Ochiai, A., Tomita, M., Esumi, H., and Soga, T. (2009) Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Res.* **69,** 4918–4925
82. Kudo, S., Rubio, C. A., Teixeira, C. R., Kashida, H., Kogure, E. et al. (2001) Pit pattern in colorectal neoplasia: endoscopic magnifying view. *Endoscopy* **33,** 367

Web link to Supplementary tables and figures (available online):

http://www.mcponline.org/content/early/2014/02/24/mcp.M113.035105/suppl/DC1

9 Acknowledgements

My profound gratitude goes to PD. Dr. Giancarlo Marra, who gave me the opportunity to fulfil my ambition for a doctorate degree in my chosen career area. Thank you for your encouragement, support and supervision all through the years of my study.

I sincerely thank all members of my thesis committee. To Prof. Josef Jiricny and Prof. Paola Picotti, your expert scientific contributions and productive discussions pertaining to my projects are truly acknowledged. I also express profound gratitude to Dr. Paolo Nanni for a constructive supervision of my projects at the Function Genomics Center, Zurich.

I express my gratitude to members of the Institute of Molecular Cancer Research and the Functional Genomics Center, Zurich for providing a scientific and friendly working atmosphere. I am grateful to Dr. Endre Laczko, Tshering Altherr, and Dr Jonas Grossmann of the FGCZ. Also, to Nathalie Selevsek and Asa Wahlander - your assistance and valuable discussions concerning my SRM project are highly appreciated.

I offer my deep appreciation to Dr. Teresa Staianao and Dr. Federico Buffoli (gastroenterologists at the Hospital of Cremona, Italy), and to the patients, whose diligent and committed attitude to scientific research facilitated the availability of samples for my studies.

To my friends, colleagues, past and present lab members, and all who have supported me thus far, thank you. My gratitude extends to the Swiss National Science Foundation for providing financial assistance for my projects. I am also truly grateful to the administrative staff of the IMCR for their steadfast support in language, official and welfare issues.

I remain forever thankful to my parents, Engr. Godwin O. L. Ogbuka (of blessed memory) and Dr. Mrs. Laurette A. R. Ogbuka, without whom there would be no "me". Thank you for nurturing me through childhood, for rousing my initial interests in the sciences and for your unending love, support and prayers. To my siblings, thank you for persistent encouragement and friendship.

Finally, I am eternally grateful to my husband, Engr. Chimaraoke O. Uzozie - for your faithful companionship, understanding, and support. And to our children, who never deprived me of bursts of laughter and smiles on those weary, dreary days.

To God be the glory.

10 Curriculum Vitae

UZOZIE Anuli Christiana

Born November 15th, 1981 in Warri, Nigeria.
Dubendorfstrasse 32, 8051 Zurich, Switzerland
Tel: +41 788 503 583
uzozie.anuli@imcr.uzh.ch

EDUCATION

2010 - 2014 (expected)

Ph.D in Cancer Biology

University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

Major research areas

- Apply untargeted and targeted proteomic techniques to investigate protein alterations during human colorectal tumorigenesis.
- Verify proteomics data with immunobased methods and investigate target protein and molecular pathways using methods in molecular and cellular biology, transcriptomics and metabolomics.
- Perform bioinformatic analysis on large mass spectrometry data.

2008 - 2010

Master of Science in Applied Biomolecular Technology

The University of Nottingham, School of Biosciences, Loughborough, LE12 5RD, England

- Graduated with **Distinction**
- **Award** - Beckman Coulter UK Prize for the School of Biosciences, University of Nottingham

Thesis: Probing the effect of cold atmospheric plasma on soft fruits, bacteria and other selected micro-organisms.

1999 - 2004

Bachelor of Technology in Biochemistry

Federal University of Technology, Owerri, P.M.B 1526, Imo State, Nigeria

- Graduated with **First Class Honours** and a cumulative GPA of **4.56 / 5.0**
- **Award** - Best graduating student, Department of Biochemistry

Thesis: Investigating the hepatoprotective and chemopreventive effects of the medicinal herb *Chromolaena Odorata.*

OTHER OCCUPATIONAL ACTIVITIES

2013 - 2014

Graduate student in Cancer Biology

Swiss Federal Institute of Technology (ETH), Institute of Food, Nutrition and Health, Laboratory of Food, Nutrition and Toxicology, Schmelzbergstrasse 9, 8092 Zurich, Switzerland.

- Planned and performed targeted proteomic experiments (Selected Reaction Monitoring) to investigate the effect of selected bioactive food compounds on the cellular induction of reductase enzymes during colorectal tumorigenesis.

2009 - 2010

Industrial Trainee

The University of Nottingham, School of Biosciences, Loughborough, LE12 5RD, England

- Assisted a project team in investigating the mechanisms of microbial destruction/inactivation induced as a result of exposure to cold atmospheric plasma.

2006 - 2008

Executive Assistant

Zenith Bank Plc, Trans Amadi, Port Harcourt, Rivers State, Nigeria

- Performed core bank operations in funds transfer unit, organized and conducted recruitment and training of junior employees.

2002 - 2003

Industrial Trainee

Shell Petroleum Development Company, Industrial Area Hospital, Medical Laboratory Unit, Delta State, Nigeria

- Assisted medical laboratory scientists in conducting clinical diagnostic tests.

2002 - 2002

Industrial Trainee

Shell Petroleum Development Company,
SPDC Environmental Health Office, Edjeba, Warri, Nigeria

- Assisted environmental health officers in collecting samples and performing routine microbiology and risk assessment tests.

2005 - 2006

Science Tutor

National Youth Service Corp (NYSC)
Government Secondary School, Borikiri, Port Harcourt, Rivers State, Nigeria

Tutored junior secondary school pupils in a major science subject.

SCIENTIFIC PUBLICATIONS AND PRESENTATIONS

Scientific Publications

1. Uzozie A., Nanni P., Grossman J., Barkow-Oesterreicher S., Staiano T., Buffoli F., Shay J. W., Tiwari, A., Laczko E., and Marra G. (2013) "SORD overexpression and other aspects of dysregulated protein expression in human precancerous colorectal neoplasms: A quantitative proteomics study" (Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.035105, 1198–1218, 2014)
2. Tiwari A., Schneider M., Fiorino A., Haider R., Okoniewski M.J., Roschitzki B., Uzozie A., Menigatti M., Jiricny J., and Marra G. (2013) "Early Insights into the Function of KIAA1199, a markedly overexpressed protein in human colorectal tumors". (PLoS ONE 8(7): e69473. doi:10.1371/journal.pone.0069473)
3. Erzinger M., Bovet, C., Uzozie A., and Sturla s. (2014) "Induction of complementary function reductase enzymes in colon cancer cells by dithiole-3-thione vs. sodium selenite". (Journal of Biochemical and Molecular Toxicology) (*Submitted*).
4. Uzozie A., Selevsek N., Nanni P., and Marra G. (2014) "An optimized large-scale SRM method for the detection and verification of early markers of colorectal tumorigenesis". (*In preparation*).

Scientific Presentations

1. "Quantitative shotgun proteomics reveals early markers of colorectal carcinogenesis". American Society for Biochemists and Molecular Biology (ASBMB), Experimental Biology Meeting, 26-30 April 2014, San Diego, California, United States. (Oral and Poster presentation)
2. "SORD overexpression and other aspects of dysregulated protein expression in human precancerous colorectal neoplasms: A quantitative proteomics study". 20th Congress of the Italian Society of Digestive Diseases, FISMAD, 19-22 March 2014, Napoli, Italy. (Oral presentation)
3. "A shotgun proteomic approach reveals significant protein expression changes in precancerous colorectal neoplasms". Digestive Disease Week, American Gastroenterological Association, 18-21 May 2013, Orlando, Florida, USA. (Oral presentation)
4. "Identification of proteome alterations characteristic of the early stages of colorectal tumorigenesis", 11th Charles Rodolphe Brupbacher Symposium, 30 January - 1 February 2013, Zurich, Switzerland. (Poster presentation)
5. "Proteomic analysis of human colorectal precancerous lesions by quantitative mass spectrometry", 11th Human Proteome Organisation (HUPO) World Congress, 9-13 September 2012, Boston, Massachusetts, USA. (Poster presentation)
6. "Proteomic analysis of human colorectal adenomatuous tumors", 4th Cancer Biology PhD Students Retreat, 7-12 March 2012, Interlaken, Switzerland. (Oral presentation)
7. Proteomic analysis of human colorectal precancerous lesions", 4th University Research Priority Program Retreat, 2-4 October 2011, Grindelwald, Switzerland. (Oral presentation)

AWARDS AND SCHOLARSHIPS

- Best Poster award, American Society for Biochemistry and Molecular Biology (ASBMB) (2014)
- Travel award, American Society of Biochemistry and Molecular Biologisty (ASBMB) (2014)
- The University of Nottingham, Developing Solutions Scholarship (2008)

LANGUAGES

English (Proficient)

German (Basic), French (Basic)